

FORM PTO-1390 (Modified)  
(REV 11-98)

U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

ATTORNEY'S DOCKET NUMBER

TRANSMITTAL LETTER TO THE UNITED STATES  
DESIGNATED/ELECTED OFFICE (DO/EO/US)  
CONCERNING A FILING UNDER 35 U.S.C. 371

112740-213

U.S. APPLICATION NO. (IF KNOWN, SEE 37 CFR

09/830497

INTERNATIONAL APPLICATION NO.  
PCT/DE00/02917INTERNATIONAL FILING DATE  
August 25, 2000PRIORITY DATE CLAIMED  
August 26, 1999

TITLE OF INVENTION

METHOD FOR TRAINING A SPEAKER RECOGNITION SYSTEM

APPLICANT(S) FOR DO/EO/US

Marcin KUROPATWINSKI

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☒ This is an express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).
4. ☐ A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. ☒ A copy of the International Application as filed (35 U.S.C. 371 (c) (2))
  - a. ☒ is transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☐ has been transmitted by the International Bureau.
  - c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. ☒ A copy of the International Search Report (PCT/ISA/210).
8. ☒ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371 (c)(3))
  - a. ☐ are transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☐ have been transmitted by the International Bureau.
  - c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
  - d. ☒ have not been made and will not be made.
9. ☐ A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
10. ☒ An oath or declaration of the inventor(s) (35 U.S.C. 371 (c)(4)).
11. ☐ A copy of the International Preliminary Examination Report (PCT/IPEA/409)
12. ☐ A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371 (c)(5)).

## Items 13 to 20 below concern document(s) or information included:

13. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
14. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
15. ☐ A **FIRST** preliminary amendment.
16. ☐ A **SECOND** or **SUBSEQUENT** preliminary amendment.
17. ☐ A substitute specification.
18. ☐ A change of power of attorney and/or address letter.
19. ☒ Certificate of Mailing by Express Mail
20. ☒ Other items or information:

Return Receipt Postcard

U.S. APPLICATION NO. (IF KNOWN, SEE 37 CFR

097830497

INTERNATIONAL APPLICATION NO

PCT/DE00/02917

ATTORNEY'S DOCKET NUMBER

112740-213

21. The following fees are submitted:

**BASIC NATIONAL FEE ( 37 CFR 1.492 (a) (1) - (5)) :**

- ☐ Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO ..... \$1,000.00
- ☒ International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO ..... \$860.00
- ☐ International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO ..... \$710.00
- ☐ International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provisions of PCT Article 33(1)-(4) ..... \$690.00
- ☐ International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4) ..... \$100.00

**ENTER APPROPRIATE BASIC FEE AMOUNT =****\$860.00**

Surcharge of \$130.00 for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492 (e)).

**\$0.00**

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE
Total claims	6 - 20 =	0	x \$18.00
Independent claims	1 - 3 =	0	x \$80.00

**\$0.00****\$0.00**Multiple Dependent Claims (check if applicable). ☐**\$0.00****TOTAL OF ABOVE CALCULATIONS =****\$860.00**

Reduction of 1/2 for filing by small entity, if applicable. Verified Small Entity Statement must also be filed (Note 37 CFR 1.9, 1.27, 1.28) (check if applicable). ☐

**\$0.00****SUBTOTAL =****\$860.00**

Processing fee of \$130.00 for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492 (f)).

**\$0.00****TOTAL NATIONAL FEE =****\$860.00**

Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31) (check if applicable). ☐

**\$0.00****TOTAL FEES ENCLOSED =****\$860.00**

Amount to be:  
refunded  
charged

\$  
\$

- ☒ A check in the amount of **\$860.00** to cover the above fees is enclosed.
- ☐ Please charge my Deposit Account No. \_\_\_\_\_ in the amount of \_\_\_\_\_ to cover the above fees.  
A duplicate copy of this sheet is enclosed.
- ☒ The Commissioner is hereby authorized to charge any fees which may be required, or credit any overpayment to Deposit Account No. **02-1818** A duplicate copy of this sheet is enclosed.

**NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.**

SEND ALL CORRESPONDENCE TO:

William E. Vaughan, Esq.  
Bell, Boyd & Lloyd LLC  
P.O. Box 1135  
Chicago, Illinois 60690-1135  
Tel: (312) 807-4292

SIGNATURE

William E. Vaughan

NAME

39,056

REGISTRATION NUMBER

April 26, 2001

DATE

BOX PCT

IN THE UNITED STATES ELECTED/DESIGNATED OFFICE  
OF THE UNITED STATES PATENT AND TRADEMARK OFFICE  
UNDER THE PATENT COOPERATION TREATY-CHAPTER I

5

**PRELIMINARY AMENDMENT**

APPLICANT: Marcin Kuropatwinski DOCKET NO: 112740-213  
SERIAL NO: 09/830,497 GROUP ART UNIT:  
EXAMINER:  
INTERNATIONAL APPLICATION NO: PCT/DE00/02917  
INTERNATIONAL FILING DATE: 25 August 2000  
INVENTION: METHOD FOR THE IDENTIFICATION OF SPEAKERS ON  
THE BASIS OF THEIR VOICES

10

15

Assistant Commissioner for Patents,  
Washington, D.C. 20231

Sir:

20

Please amend the above-identified International Application before entry into  
the National stage before the U.S. Patent and Trademark Office under 35 U.S.C. §371  
as follows:

**In the Specification:**

Please replace the Specification of the present application, including the  
Abstract, with the following Substitute Specification:

25

**S P E C I F I C A T I O N**

**TITLE**

**METHOD FOR THE IDENTIFICATION OF SPEAKERS ON THE BASIS  
OF THEIR VOICES**

30

**BACKGROUND OF THE INVENTION**

**Field of the Invention**

The present invention generally relates to a method for the identification of speakers on the basis of their voices, wherein the method is robust, safe, secure and reliable.

### **Description of the Prior Art**

5           The problem of speaker identification is to distinguish between different speakers or to check the predetermined speaker identity, with the only input information being the recording of the voice of the speaker.

Problems have developed relating to the access system being outwitted when the voice and the keyword are recorded by third parties.

10           Moreover, when complex probability distributions for the speech parameters of a speaker are stored, a compromise must be made between accuracy and memory requirement. For this reason, methods for storage of the probability distributions have been proposed which can be used as a function of the number of speakers.

15           Until now, the speaker has been identified, for example, with the aid of hidden Markov models or by vector quantization, see reference [1].

          The speech signal parameters used in the past, such as Cepstral AR parameters, do not provide a satisfactory solution to speaker identification problems. For this reason, other parameters need to be used, such as parameters  
20 relating to the excitation of the vocal tract, which include information that is dependent on the speaker and is at the same time largely phoneme-independent.

### **SUMMARY OF THE INVENTION**

          In light of the above, the present invention provides a method for estimation of the probability distribution of the coder parameters for the respective speaker, as  
25 well as a method which prevents the access system from being outwitted. In addition, the present invention solves the problem of speaker identification based on the parameters of an analysis via synthesis coders using linear prediction (LPAS) [1] (for example, a harmonic vector excited codec [5] or waveform interpolation codec [4].

Accordingly, the present invention is first directed to a method for identifying speakers on the basis of their voices, wherein the method includes a preparation phase, a simulated usage phase of the training phase, and a usage phase. In the preparation phase, the method includes the steps of: segmenting into first  
5 speech signal frames of a given length, a number of text-dependent or text-independent reference spoken expressions, from a number of speakers, which form a speaker-related training statement; supplying the first speech signal frames to an analysis-by-synthesis coder based on linear predictions; calculating a first short-term predictor parameter, a first long-term predictor parameter and/or an excitation  
10 parameter for the coder, in the analysis-by-synthesis coder for each of the number of speakers and for each first speech signal frame, wherein the parameters form speaker-related training material; calculating the frequency of the respective occurrence of the first parameters in the speaker-related training statement and/or the probability densities with which the first parameters are contained in the  
15 speaker-related training statement, in the analysis-by-synthesis coder for each of the number of speakers and for each first speech signal frame; and storing the calculated frequencies and/or the probability densities on a speaker-related basis as speaker data. In the simulated usage phase the method includes the steps of: segmenting into second speech signal frames of a given length L a text-dependent  
20 or a text-independent simulation spoken expression of a given speaker; supplying the second speech signal frames to the analysis-by-synthesis coder; calculating a second short-term predictor parameter, a second long-term predictor parameter and/or a second excitation parameter for the coder, the calculation being performed in the analysis-by-synthesis coder for the given speaker and for every other speech  
25 signal frame in each case; calculating first probability hits for every other speech signal frame from the calculated second parameters and the speaker data stored for the given speaker in the preparation phase, the probability hits indicating a probability with which the second parameters match the first parameters; combining the first probability scores from all the second speech signal frames; and  
30 checking to determine whether the combined first probability scores are greater

than a predetermined first threshold which confirms the voice of the given speaker, when the combined first probability scores are greater than the predetermined first threshold, the voice of the given speaker is confirmed, and when the combined first probability scores are less than or equal to the predetermined first threshold, the preparation phase continues for further reference spoken expressions by the given speaker until the voice of the given speaker is confirmed. In the usage phase, the method includes the steps of: segmenting a text-dependent or a text-independent used spoken expression of the given speaker into third speech signal frames of a given length; supplying the third speech signal frames to the analysis-by-synthesis coder; calculating a third short-term predictor parameter, a third long-term predictor parameter and/or a third excitation parameter for the coder, in the analysis-by-synthesis coder for the given speaker and for every third speech signal frame in each case; calculating second probability hits for every third speech signal frame from the calculated third parameters and the speaker data stored for the given speaker in the preparation phase, the second probability hits indicating a probability with which the third parameters have been spoken by the given speaker; combining the second probability hits from all the third speech signal frames; and checking to determine whether the combined second probability scores are greater than a predetermined second threshold and the voice of the given speaker, when the combined second probability hits are greater than the predetermined second threshold, the voice of the given speaker is identified, and where the combined second probability scores are less than or equal to the predetermined second threshold, the voice of the given speaker is not identified.

In an embodiment of the method, a harmonic vector excited predictive coder and a waveform interpolating coder is used as a parametric coder.

In an embodiment of the method, an LPAS coder is used as the analysis-by-synthesis coder.

In an embodiment, the method further includes the step of quantizing the frequencies and/or the probability densities using a vector quantizer having a specific and considerably reduced number of bits.

In an embodiment, the method further includes the step of entering noise which is known to the speaker identification system when the spoken expression of a speaker is entered into the speaker identification system.

5 In an embodiment, the method further includes the step of subtracting the entered noise internally, before the segmentation, from the recording of the speakers voice.

Additional features and advantages of the present invention are described in, and will be apparent from, the following Detailed Description of the Preferred Embodiments and the Drawings.

## 10 DESCRIPTION OF THE DRAWINGS

Figure 1 shows, in generally diagrammatic form, the object of speaker identification;

Figure 2 is a general flowchart of the basic considerations in speaker verification;

15 Figure 3 schematically illustrates the synthesis model of a CELP coder;

Figure 4 shows a schematic illustration of the various parameter groups of an LPAS coder;

Figure 5 shows a block diagram form speaker identification using the parameters of an LPAS coder;

20 Figure 6 shows a block diagram form speaker identification using the parameters of an LPAS coder, wherein probability densities are stored together with the code vectors for the parameters;

Figure 7 shows in detailed flowchart form speaker verification using the parameters of an LPAS coder; and

25 Figures 8a-8m show in flowchart form an exemplary embodiment of all phases of the method of the present invention.

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

### Speaker identification

30 In systems for *speaker identification*, statistical principles [2] are used to check whether the spoken sentence has been spoken by one of the speakers covered

by the speaker identification system. In the process, there are, in principle, two types of speaker identification systems; text-dependent systems and text-independent systems. For the procedure described in the present invention, text independence of the system is achieved via an expanded training phase in which the speaker has to record a wide range of material, and the probability distributions of the speech signal parameters are established from all the spoken material. A text-dependent system can be trained more easily since the spoken material which is spoken by the speaker during the usage phase is limited to a number of keywords or specific sentences. The preparation phase is continued until the system reliably identifies the voice of the speaker. An exemplary embodiment of all phases of the method of the present invention is summarily depicted and described in Figures 8a-8m, to which the entirety of the following disclosure is associated.

The object of *speaker identification* is illustrated in Figure 1 (Problem of speaker identification).

*Speaker identification* is dealt with as a problem relating to the detection of multiples [2]. The classes to be distinguished between, one for each speaker who is intended to be identified by the system, are referred to as  $sp_i = 1..M$ , where  $M$  is the number of speakers covered by the speaker identification system. Speaker identification is based on recorded signals spoken by the respective speaker. The speech signal is segmented into the signal frames  $x = [x(1)..x(K)]$  ( $K = 160$  for a signal frame with a length of 20 ms and a sampling frequency of 8 kHz, for example). The segmentation process produces the speech signal frames  $x(1)..x(N)$ , where  $N$  depends on the total length of the sentence or keyword spoken by the speaker. The decision on the speaker is made from the probabilities or probability densities (referred to jointly as probability scores) that the vectors of the samples  $x(l)$   $l = 1..N$  belong to the class  $sp_i$ . The statistically optimum decision scheme selects that class  $sp_i$  having the highest probability value for given  $x(l)$ ,  $l = 1..N$ . As such, the vector  $x(l)$  is assigned to the class  $sp_j$  for which:



$$p(\mathbf{x}(1)...\mathbf{x}(N) | sp_j) > p(\mathbf{x}(1)...\mathbf{x}(N) | sp_i) \text{ for all } j \neq i$$

### Speaker verification

5 The problem of *speaker verification* is to check the predetermined identity of the speaker on the basis of his/her voice. This corresponds to the situation illustrated in Figure 2 (Problem of speaker verification).

The process of speaker verification is carried out in a similar manner to that of speaker identification, wherein the spoken sentence is likewise segmented. However, after this, the voice is not classified, but a probability score is calculated  
10 for the predetermined speaker identity and is compared with a threshold. The identity of the speaker is thus confirmed on the basis of his/her voice when:

$$p(\mathbf{x}(1)...\mathbf{x}(N) | sp_j) > \text{threshold}$$

where  $sp_j$  corresponds to the predetermined speaker identity. The threshold must be set sufficiently high to avoid the situation in which a speaker with a different  
15 identity to that predetermined is accepted/authorized.

### LPAS coder

The speech coding methods used nowadays are predominantly based on the analysis by synthesis method using an LPC synthesis filter [2]. In these methods, speech coding is optimized by repetition of the coding and decoding operations  
20 until the optimum parameter set is found for the given speech section.

One of the most widely used types of LPAS coder is the CELP coder. One relatively new development is the harmonic vector excited codec, where the form of the excitation signals is particularly suitable for the described task. Figure 3 (Design of an LPAS Coder) illustrates the synthesis model of a CELP coder. The  
25 synthesis model defines the method for calculating the synthesized speech signal from the quantized parameters of the speech signal. In general, each LPAS coder has the following parameter groups (see also Figure 4):

- Short-term predictor parameters. The short-term predictor parameters are generally calculated via classical LPC analysis, using the correlation method or the covariance method for linear prediction [3]. 8-10 LPC coefficients are used for signal frames with a length of 20 to 30 ms and a  
5 sampling rate of 8 kHz. The short-term predictor parameters may occur in various forms (for example the reflection coefficients or in the form of line spectrum frequencies LSF), depending on the representation which can be quantized better. It has been found that the LSF coefficients are most suitable for quantization, and this form of prediction coefficients is generally used. The  
10 short-term predictor parameters are calculated using an open-loop procedure, that is to say without the overall optimization, illustrated in Figure 1, with the other parameters relating to the synthesis error.
- Long-term predictor parameters. Long-term predictor parameters are used in a filter which synthesizes the fundamental frequency of the speech signal. This is  
15 generally a long-term predictor with a filter coefficient and a parameter for the fundamental period of the voice signal. A long-term predictor with the parameters  $b = [b, N]$  is a part of Figure 2. The long-term predictor parameters are likewise calculated using an open-loop procedure, without overall optimization with the other parameters. In some coders, a refined search is  
20 sometimes carried out based on the long-term predictor parameters using a closed-loop procedure.
- The excitation parameters. The 5-10 ms subframes of the remaining signal are vector-quantized using a closed-loop procedure in a CELP coder. The transmitted parameters allow the signal forms to be reproduced at the decoder  
25 end from the stored codebook.

In an HVXC codec, the output from the LPC analysis filter is transformed to the frequency domain and the fundamental-period-normalized spectral envelope is vector-quantized.

Speaker identification using the parameters of an LPAS coder

The speech coder parameters provide a comprehensive description of the possible speech signals using considerably fewer parameters than when the speech signal is represented as a sequence of samples.

5 The decomposition of the speech signal into parameter groups can be used in various ways for speaker identification. The methods for calculation of the parameters and synthesis of the speech signal imply probability density estimation methods (for example the probabilities of the parameters, which are regarded as discrete probability variables). Those defined using a closed-loop procedure should actually be regarded as discrete probability variables, since it is impossible to link  
10 the volumes of the parameter space regions of the vector quantizer for parameters such as these. This relates in particular to the excitation parameters. The probability distributions for such parameters are estimated by calculating relative frequencies of the parameters/code vectors in the training statement.

Those which are calculated using an open-loop procedure in the coder are  
15 initially available in a non-quantized form and are quantized only after this, with vector quantization generally being used. For parameters such as these, the probability densities can be estimated from the training statement. This approach is used primarily for the short-term predictor parameters.

The probability density estimation is based on the histogram method [6].  
20 This method requires knowledge of the volumes of those regions of the parameter space which are linked to the quantized points.

A method for storage of probability distributions is obtained, according to Figure 5 (Speaker identification using the parameters of an LPAS coder), if the possible code vectors for the speech signal parameters are stored once for the entire  
25 population, which corresponds to the situation where the quantization steps/code vectors are determined once, from the database which contains the recordings by a large number of speakers. The probability distributions of the parameters for the speakers are then stored together with the indices of the code vectors for the parameters in the system. This is suitable for large systems with a very large

number of users (ATM, access systems in companies). In this regard, see all Figure 7.

Another method is for the code vectors for the parameters for each speaker to be trained individually. The code vectors are then stored together with the values of the probability densities at those parameter space points defined by the code vectors. One possible way of carrying out this method is shown in Figure 6 (speaker identification using the parameters of an LPAS coder, probability densities are stored together with the code vectors for the parameters). This method is intended for a small number of speakers (for example, for a voice-controlled door in a dwelling).

#### Training phase of a speaker identification system

The probability density distributions for the speaker classes are estimated from the training material. For text-dependent speaker identification (speaker identification/speaker verification), a specific sentence or keyword is repeated during the training phase until the speaker identification operates reliably.

Phonetically balanced spoken material must be recorded for text-independent speaker verification. In this case as well, the training phase must be repeated until the speaker identification/verification operates reliably.

The material recorded during the training phase is in each case used with a phase shift a number of times for training, in order to make the speaker identification system independent of the initial phase of the recorded voices. The data used for training are referred to as the training statement  $TS_{sp_i}$ , with  $sp_i$  symbolizing the speaker.

#### *Estimation of the probability densities*

In order to describe the method according to the present invention for estimation of the probability densities of the parameters for the speaker classes, a number of necessary definitions first will be introduced.

The introduced abstraction of the coding process has the advantage that the estimation of the probability densities can be described in a simple manner without needing to go into the details of the highly complicated operations in the speech

coder. A detailed description of the parameter calculation process can be found in [4] and [5].

A speech coder operates in evaluation intervals. The operations described in that section via the LPAS coder are carried out in the speech coder for each signal frame, and supply the parameters of the speech signal for the respective frame.

Calculation of a non-quantized parameter vector  $p$  from the signal frame  $x$  is written as  $p = K_p(x)$  in an open-loop optimization procedure. The quantization of the parameter is referred to as  $\hat{p} = Q_p(p)$ . That region in the parameter space of the parameter  $p$  which is mapped onto the code vector  $\hat{p}$  in the coding process is referred to as  $S_{\hat{p}} = \{p : Q_p(p) = \hat{p}\}$ . The volume of this region is referred to as  $V(S_{\hat{p}})$ .

The set of possible code vectors for the parameter  $p$  is written as  $C_p = \{\hat{p}_i; i=1..N_p\}$ , where  $N_p$  is the number of code vectors. The set or regions which are linked to the code vectors is referred to as  $R_p = \{S_i; i=1..N_p\}$ . The association function for a region  $S_i$  is referred to as:

$$1_{S_i}(p) = \begin{cases} 1 & \text{for } p \in S_i \\ 0 & \text{for } p \notin S_i \end{cases}$$

The frequency of occurrence of a parameter in the training statement is calculated using

$$f_{S_i} = \frac{\text{Number of parameter values from the training statement } TS_{sp_i} \text{ which occur in the region } S_i}{\text{Number of parameter values from the training statement } TS_{sp_i}}$$

The estimated probability density distribution then becomes:

$$p(p | sp_i) = \sum_{k=1}^{N_p} 1_{S_k}(p) \frac{f_{S_k}}{V(S_k)}$$

*Estimation of the probabilities*

The probability functions (probability mass functions) are estimated for those parameters which are regarded as discrete probability variables, that is to say in particular the excitation from the codebook, which is optimized using a closed-loop procedure, and the fundamental period of the speech signal. These probability functions are defined as the frequencies of the given parameter codes in the training statement for the respective speaker.

#### *Storage of the probability distributions*

The speech parameters are not all calculated at the same time, but successively, in a speech coder. For example, the short-term predictor parameters are calculated first, and the remaining parameters for already known short-term predictor parameters are then optimized with regard to the synthesis or the prediction error. This allows effective storage of the probability distributions as conditional probabilities of the code vectors in a tree structure. This is possible due to the following relationship:

$$p(p_K, p_L, p_A | sp_i) = p(p_K | sp_i) p(p_L | sp_i, p_K) p(p_A | sp_i, p_K, p_L)$$

$p_K$  - Vector for a short-term parameter  
 $p_L$  - Vector for a long-term parameter  
 $p_A$  - Vector for an excitation parameter

A major simplification can be achieved if the speech parameters within a signal frame can be assumed to be statistically independent. The above formula then becomes:

$$p(p_K, p_L, p_A | sp_i) = p(p_K | sp_i) p(p_L | sp_i) p(p_A | sp_i)$$

The probability densities need to be stored at a very large number of points in parameter space in the system. The number of bits used for storing probability densities is critical to the complexity of the overall system. A vector quantizer is therefore used for the probability values. This makes it possible to reduce the number of bits used for storing the probability distributions.

#### *System safety and security*

In order to prevent the system from being outwitted, noise is transmitted at the same time that the voice of the speaker is being recorded, which noise is known to the system, and from which the digitized speech signal is subtracted.

The present invention can be used for access control applications, such as voice-controlled doors, or for verification; for example, for bank access systems. The procedure can be implemented as a program module on a processor which carries out the task of speaker identification in the system.

Although the present invention has been described with reference to specific embodiments, those of skill in the art will recognize that changes may be made thereto without departing from the spirit and scope of the invention as set forth in the hereafter appended claims.

- [1] S. Furui, "Recent advances in speaker recognition", Pattern Recognition Letters, Tokyo Inst. of Technol., 1997
- [2] P. Vary, U. Heute, W. Hess, *Digitale Sprachsignalverarbeitung [Digital speech signal processing]*, B.G. Teubner, Stuttgart, 1998
- [3] K. Kroschel, *Statistische Nachrichtentheorie [Statistical information theory]*, 3rd ed., Springer-Verlag, 1997
- [4] W.B. Kleijn, K.K. Paliwal, *Speech Coding and Synthesis*, Elsevier, 1995
- [5] ISO/IEC 14496-3, MPGA-3 HVXC Speech Coder description
- [6] Prakasa Rao, *Functional Estimation*, Academic Press, 1982

#### **ABSTRACT OF THE DISCLOSURE**

A method for speaker identification using parameters of an LPAS coder or of a parametric coder for modeling the probability distribution for the speaker classes.

**In the claims:**

On page 12, cancel line 1, and substitute the following left-hand justified heading therefor:

**I Claim as My Invention:**

5           Please cancel claims 1-6, without prejudice, and substitute the following claims therefor:

7.       A method for the identification of speakers on the basis of the speakers' respective voices, the method comprising the steps of:

10           segmenting, in a preparation phase, into first speech signal frames of a given length, a plurality of one of text-dependent and text-independent reference spoken expressions, from a plurality of speakers, which form a speaker-related training statement;

          supplying the first speech signal frames, in the preparation phase, to an analysis-by-synthesis coder based on linear predictions;

15           calculating, in the preparation phase, at least one of a first short-term predictor parameter, a first long-term predictor parameter and a first excitation parameter for the coder in the analysis-by-synthesis coder for each of the plurality of speakers and for each first speech signal frame in each case, wherein the parameters form speaker-related training material;

20           calculating, in the preparation phase, at least one of a frequency of a respective occurrence of the first parameters in the speaker-related training statement and probability densities with which the first parameters are contained in the speaker-related training statement, the calculation being performed in the analysis-by-synthesis coder for each of the plurality of speakers and for each first  
25   speech signal frame in each case;

          storing, in the preparation phase, at least one of the calculating frequencies and the probability densities on a speaker-related basis as speaker data;

          segmenting, in a simulated usage phase of the training phase, into second speech signal frames of a given length, one of a text-dependent and a text-  
30   independent simulation spoken expression of a given speaker;



supplying, in the simulated usage phase, the second speech signal frames to the signal-by-synthesis coder;

calculating, in the simulated usage phase, at least one of a second short-term predictor parameter, a second long-term predictor parameter and a second  
5 excitation parameter for the coder, the calculation being performed in the analysis-by-synthesis coder for the given speaker and for every other speech signal frame in each case;

calculating, in the simulated usage phase, first probability hits for every other speech signal frame from the calculated second parameters and the speaker  
10 data stored for the given speaker in the preparation phase, the probability hits indicating a probability with which the second parameters match the first parameters;

combining, in the simulated usage phase, the first probability scores from all the second speech signal frames;

15 checking, in the simulated usage phase, to determine whether the combined first probability scores are greater than a predetermined first threshold which confirms the voice of the given speaker, when the combined first probability scores are greater than the predetermined first threshold, the voice of the given speaker is confirmed, and when the combined first probability scores are less than or equal to  
20 the predetermined first threshold, the preparation phase continues for further reference spoken expressions by the given speaker until the voice of the given speaker is confirmed;

segmenting into third speech signal frames of a given length, in a usage phase, one of a text-dependent and a text-independent used spoken expression of  
25 the given speaker;

supplying, in the usage phase, the third speech signal frames to be analysis-by-synthesis coder;

calculating, in the usage phase, at least one of a third short-term predictor parameter, a third long-term predictor parameter and a third excitation parameter

for the coder, the calculation being performed in the analysis-by-synthesis coder for the given speaker and for every third speech signal frame in each case;

calculating, in the usage phase, second probability hits for every third speech signal frame from the calculated third parameters and the speaker data stored for the given speaker in the preparation phase, the second probability hits indicating a probability with which the third parameters have been spoken by the given speaker;

combining, in the usage phase, the second probability hits from all the third speech signal frames; and

checking, in the usage phase, to determine whether the combined second probability scores are greater than a predetermined second threshold which identifies the voice of the given speaker, when the combined second probability hits are greater than the predetermined second threshold, the voice of the given speaker is identified, and when the combined second probability scores are less than or equal to the predetermined second threshold, the voice of the given speaker is not identified.

8. A method for the identification of speakers on the basis of the speakers respective voices as claimed in claim 7, wherein one of a harmonic vector excited predictive coder and a waveform interpolating coder is used as a parametric coder.

9. A method for the identification of speakers on the basis of the speakers respective voices as claimed in claim 7, wherein an LPAS coder is used as the analysis-by-synthesis coder.

10. A method for the identification of speakers on the basis of the speakers respective voices as claimed in claim 7, the method further comprising the step of:

quantizing at least one of the frequencies and the probability densities using a vector quantizer having a specific and considerably reduced number of bits.

11. A method for the identification of speakers on the basis of the  
5 speakers respective voices as claimed in claim 7, the method further comprising the step of:

entering noise which is known to the speaker identification system when the spoken expression of a speaker is entered into the speaker identification system.

10 12. A method for the identification of speakers on the basis of the speakers respective voices as claimed in claim 11, the method further comprising the step of:

subtracting the entered noise internally, before the segmentation, from the recording of the speakers voice.

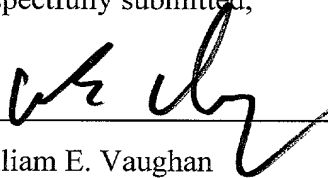
15 **REMARKS**

The present amendment makes editorial changes and corrects typographical errors in the specification, which includes the Abstract, in order to conform the specification to the requirements of United States Patent Practice. No new matter is added thereby. Attached hereto is a marked-up version of the changes made to the  
20 specification by the present amendment. The attached page is captioned "**Version With Markings To Show Changes Made**".

In addition, the present amendment cancels original claims 1-6 in favor of new claims 7-12. Claims 7-12 have been presented solely because the revisions by red-lining and underlining which would have been necessary in claims 1-6 in order  
25 to present those claims in accordance with preferred United States Patent Practice would have been too extensive, and thus would have been too burdensome. The present amendment is intended for clarification purposes only and not for substantial reasons related to patentability pursuant to 35 USC §§103, 102, 103 or 112. Indeed, the cancellation of claims 1-6 does not constitute an intent on the part  
30 of the Applicant to surrender any of the subject matter of claims 1-6.

Early consideration on the merits is respectfully requested.

Respectfully submitted,



(Reg. No. 39,056)

William E. Vaughan  
Bell, Boyd & Lloyd LLC  
P.O. Box 1135  
Chicago, Illinois 60690-1135  
(312) 807-4292  
Attorneys for Applicant

## VERSIONS WITH MARKINGS TO SHOW CHANGES MADE

### In The Specification:

The Specification of the present application, including the Abstract, has been amended as follows:

#### SPECIFICATION

##### TITLE

~~Method for identification of speakers on the basis of their voices~~

#### METHOD FOR THE IDENTIFICATION OF SPEAKERS ON THE BASIS OF THEIR VOICES

##### BACKGROUND OF THE INVENTION

##### Description

##### Field of the Invention

~~The invention relates to a method for identification of speakers on the basis of their voices.~~

~~The object on which the invention is based is to specify a method for identification of speakers on the basis of their voices, which method is robust, safe, secure and reliable.~~

~~According to the invention, this object is achieved by the features specified in patent claim 1.~~

~~The invention will be described in more detail in the following text using a flowchart.~~

~~1.~~

The present invention generally relates to a method for the identification of speakers on the basis of their voices, wherein the method is robust, safe, secure and reliable.

##### Description of the Prior Art

~~The invention allows the identification of the speaker on the basis of his voice.~~ The problem of speaker identification is to distinguish between different speakers or to check the predetermined speaker identity, with the only input information being the recording of the voice of the speaker.

Furthermore, a method is proposed which prevents Problems have developed relating to the access system from being outwitted when the voice and the keyword are recorded by third parties.

When Moreover, when complex probability distributions for the speech parameters of a speaker are stored, a compromise must be made between accuracy and memory requirement. For this reason, methods for storage of the probability distributions have been proposed which can be used as a function of the number of speakers.

2.

Until now, the speaker has been identified, for example, with the aid of hidden Markov models or by vector quantization, see reference [1].

3.

The invention solves the problem of speaker identification based on the parameters of an analysis by means of synthesis coders using linear prediction (LPAS) [1] (for example a harmonic vector excited codec [5] or waveform interpolation codec [4]). The speech signal parameters used in the past, such as Cepstral AR parameters, do not provide a satisfactory solution to the problem speaker identification problems. For this reason, other parameters need to be used, such as parameters relating to the excitation of the vocal tract, which include information that is dependent on the speaker and is at the same time largely phoneme-independent.

### SUMMARY OF THE INVENTION

Furthermore, the In light of the above, the present invention provides a method for estimation of the probability distribution of the coder parameters for the respective speaker is given, as well as a method which prevents the access system from being outwitted. In addition, the present invention solves the problem of speaker identification based on the parameters of an analysis via synthesis coders using linear prediction (LPAS) [1] (for example, a harmonic vector excited codec [5] or waveform interpolation codec [4].

Accordingly, the present invention is first directed to a method for identifying speakers on the basis of their voices, wherein the method includes a preparation phase, a simulated usage phase of the training phase, and a usage phase. In the preparation phase, the method includes the steps of: segmenting into first speech signal frames of a given length, a number of text-dependent or text-independent reference spoken expressions, from a number of speakers, which form a speaker-related training statement; supplying the first speech signal frames to an analysis-by-synthesis coder based on linear predictions; calculating a first short-term predictor parameter, a first long-term predictor parameter and/or an excitation parameter for the coder, in the analysis-by-synthesis coder for each of the number of speakers and for each first speech signal frame, wherein the parameters form speaker-related training material; calculating the frequency of the respective occurrence of the first parameters in the speaker-related training statement and/or the probability densities with which the first parameters are contained in the speaker-related training statement, in the analysis-by-synthesis coder for each of the number of speakers and for each first speech signal frame; and storing the calculated frequencies and/or the probability densities on a speaker-related basis as speaker data. In the simulated usage phase the method includes the steps of: segmenting into second speech signal frames of a given length L a text-dependent or a text-independent simulation spoken expression of a given speaker; supplying the second speech signal frames to the analysis-by-synthesis coder; calculating a second short-term predictor parameter, a second long-term predictor parameter and/or a second excitation parameter for the coder, the calculation being performed in the analysis-by-synthesis coder for the given speaker and for every other speech signal frame in each case; calculating first probability hits for every other speech signal frame from the calculated second parameters and the speaker data stored for the given speaker in the preparation phase, the probability hits indicating a probability with which the second parameters match the first parameters; combining the first probability scores from all the second speech signal frames; and checking to determine whether the combined first probability scores are greater

than a predetermined first threshold which confirms the voice of the given speaker, when the combined first probability scores are greater than the predetermined first threshold, the voice of the given speaker is confirmed, and when the combined first probability scores are less than or equal to the predetermined first threshold, the preparation phase continues for further reference spoken expressions by the given speaker until the voice of the given speaker is confirmed. In the usage phase, the method includes the steps of: segmenting a text-dependent or a text-independent used spoken expression of the given speaker into third speech signal frames of a given length; supplying the third speech signal frames to the analysis-by-synthesis coder; calculating a third short-term predictor parameter, a third long-term predictor parameter and/or a third excitation parameter for the coder, in the analysis-by-synthesis coder for the given speaker and for every third speech signal frame in each case; calculating second probability hits for every third speech signal frame from the calculated third parameters and the speaker data stored for the given speaker in the preparation phase, the second probability hits indicating a probability with which the third parameters have been spoken by the given speaker; combining the second probability hits from all the third speech signal frames; and checking to determine whether the combined second probability scores are greater than a predetermined second threshold and the voice of the given speaker, when the combined second probability hits are greater than the predetermined second threshold, the voice of the given speaker is identified, and where the combined second probability scores are less than or equal to the predetermined second threshold, the voice of the given speaker is not identified.

In an embodiment of the method, a harmonic vector excited predictive coder and a waveform interpolating coder is used as a parametric coder.

In an embodiment of the method, an LPAS coder is used as the analysis-by-synthesis coder.

In an embodiment, the method further includes the step of quantizing the frequencies and/or the probability densities using a vector quantizer having a specific and considerably reduced number of bits.



In an embodiment, the method further includes the step of entering noise which is known to the speaker identification system when the spoken expression of a speaker is entered into the speaker identification system.

5 In an embodiment, the method further includes the step of subtracting the entered noise internally, before the segmentation, from the recording of the speakers voice.

Additional features and advantages of the present invention are described in, and will be apparent from, the following Detailed Description of the Preferred Embodiments and the Drawings.

## 10 DESCRIPTION OF THE DRAWINGS

Figure 1 shows, in generally diagrammatic form, the object of speaker identification:

Figure 2 is a general flowchart of the basic considerations in speaker verification:

15 Figure 3 schematically illustrates the synthesis model of a CELP coder;

Figure 4 shows a schematic illustration of the various parameter groups of an LPAS coder:

Figure 5 shows in block diagram form speaker identification using the parameters of an LPAS coder;

20 Figure 6 shows in block diagram form speaker identification using the parameters of an LPAS coder, wherein probability densities are stored together with the code vectors for the parameters;

Figure 7 shows in detailed flowchart form speaker verification using the parameters of an LPAS coder; and

25 Figures 8a-8m show in flowchart form an exemplary embodiment of all phases of the method of the present invention.

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

### Speaker identification

30 In systems for *speaker identification*, statistical principles [2] are used to check whether the spoken sentence has been spoken by one of the speakers covered

by the speaker identification system. In the process, there are, in principle, two types of speaker identification systems; text-dependent systems and text-independent systems. For the procedure described in the present invention, text independence of the system is achieved ~~by means of~~ via an expanded training phase; in which the speaker has to record a wide range of material, and the probability distributions of ~~said~~ the speech signal parameters are established from all the spoken material. A text-dependent system can be trained more easily since the spoken material which is spoken by the speaker during the usage phase is limited to a number of keywords or specific sentences. The preparation phase is continued until the system reliably identifies the voice of the speaker. An exemplary embodiment of all phases of the method of the present invention is summarily depicted and described in Figures 8a-8m, to which the entirety of the following disclosure is associated.

The object of *speaker identification* is illustrated in Figure 1 (Problem of speaker identification).

*Speaker identification* is dealt with as a problem relating to the detection of multiples [2]. The classes to be distinguished between, one for each speaker who is intended to be identified by the system, are referred to as  $sp_i = 1..M$ , where  $M$  is the number of speakers covered by the speaker identification system. Speaker identification is based on recorded signals spoken by the respective speaker. The speech signal is segmented into the signal frames  $x = [x(1)..x(K)]$  ( $K = 160$  for a signal frame with a length of 20 ms and a sampling frequency of 8 kHz, for example). The segmentation process produces the speech signal frames  $x(1)..x(N)$ , where  $N$  depends on the total length of the sentence or keyword spoken by the speaker. The decision on the speaker is made from the probabilities or probability densities (referred to jointly as probability scores) that the vectors of the samples  $x(l)$   $l = 1..N$  belong to the class  $sp_i$ . The statistically optimum decision scheme selects that class  $sp_i$  having the highest probability value for given  $x(l)$ ,  $l = 1..N$ . ~~This means that~~ As such, the vector  $x(l)$  is assigned to the class  $sp_j$  for which:

$$p(\mathbf{x}(1)...\mathbf{x}(N) | sp_j) > p(\mathbf{x}(1)...\mathbf{x}(N) | sp_i) \text{ for all } j \neq i$$

### Speaker verification

5 The problem of *speaker verification* is to check the predetermined identity of the speaker on the basis of his/her voice. This corresponds to the situation illustrated in Figure 2 (Problem of speaker verification).

The process of speaker verification is carried out in a similar manner to that of speaker identification, ~~that is to say~~ wherein the spoken sentence is likewise segmented. However, after this, the voice is not classified, but a probability score is  
10 calculated for the predetermined speaker identity, and is compared with a threshold. The identity of the speaker is thus confirmed on the basis of his/her voice when:

$$p(\mathbf{x}(1)...\mathbf{x}(N) | sp_j) > \text{threshold}$$

where  $sp_j$  corresponds to the predetermined speaker identity. The threshold must be set sufficiently high to avoid the situation in which a speaker with a different  
15 identity to that predetermined is accepted/authorized.

### LPAS coder

The speech coding methods used nowadays are predominantly based on the analysis by synthesis method using an LPC synthesis filter [2]. In these methods, speech coding is optimized by repetition of the coding and decoding operations  
20 until the optimum parameter set is found for the given speech section.

One of the most widely used types of LPAS coder is the CELP coder. One relatively new development is the harmonic vector excited codec, where the form of the excitation signals is particularly suitable for the described task. Figure 3 (Design of an LPAS ~~coder~~ Coder) illustrates the synthesis model of a CELP coder.  
25 The synthesis model defines the method for calculating the synthesized speech signal from the quantized parameters of the speech signal. In general, each LPAS coder has the following parameter groups (see also Figure 4):

- Short-term predictor parameters. The short-term predictor parameters are generally calculated ~~by means of~~ via classical LPC analysis, using the correlation method or the covariance method for linear prediction [3]. 8-10 LPC coefficients are used for signal frames with a length of 20 to 30 ms and a sampling rate of 8 kHz. The short-term predictor parameters may occur in various forms (for example the reflection coefficients or in the form of line spectrum frequencies LSF), depending on the representation which can be quantized better. It has been found that the LSF coefficients are most suitable for quantization, and this form of prediction coefficients is generally used. The short-term predictor parameters are calculated using an open-loop procedure, that is to say without the overall optimization, illustrated in Figure 1, with the other parameters relating to the synthesis error.
- Long-term predictor parameters. Long-term predictor parameters are used in a filter which synthesizes the fundamental frequency of the speech signal. This is generally a long-term predictor with a filter coefficient and a parameter for the fundamental period of the voice signal. A long-term predictor with the parameters  $b = [b, N]$  is a part of Figure 2. The long-term predictor parameters are likewise calculated using an open-loop procedure, without overall optimization with the other parameters. In some coders, a refined search is sometimes carried out based on the long-term predictor parameters using a closed-loop procedure.
- The excitation parameters. The 5-10 ms subframes of the remaining signal are vector-quantized using a closed-loop procedure in a CELP coder. The transmitted parameters allow the signal forms to be reproduced at the decoder end from the stored codebook.

In an HVXC codec, the output from the LPC analysis filter is transformed to the frequency domain and the fundamental-period-normalized spectral envelope is vector-quantized.

Speaker identification using the parameters of an LPAS coder

The speech coder parameters provide a comprehensive description of the possible speech signals using considerably fewer parameters than when the speech signal is represented as a sequence of samples.

5 The decomposition of the speech signal into ~~the said~~ parameter groups can be used in various ways for speaker identification. The methods for calculation of the parameters and synthesis of the speech signal imply probability density estimation methods (for example the probabilities of the parameters, which are regarded as discrete probability variables). Those defined using a closed-loop procedure should actually be regarded as discrete probability variables, since it is  
10 impossible to link the volumes of the parameter space regions of the vector quantizer for parameters such as these. This relates in particular to the excitation parameters. The probability distributions for such parameters are estimated by calculating relative frequencies of the parameters/code vectors in the training statement.

15 Those which are calculated using an open-loop procedure in the coder are initially available in a non-quantized form and are quantized only after this, with vector quantization generally being used. For parameters such as these, the probability densities can be estimated from the training statement. This approach is used primarily for the short-term predictor parameters.

20 The probability density estimation is based on the histogram method [6]. This method requires knowledge of the volumes of those regions of the parameter space which are linked to the quantized points.

A method for storage of probability distributions is obtained, according to Figure 5 (Speaker identification using the parameters of an LPAS coder), if the  
25 possible code vectors for the speech signal parameters are stored once for the entire population, which corresponds to the situation where the quantization steps/code vectors are determined once, from the database which contains the recordings by a large number of speakers. The probability distributions of the parameters for the speakers are then stored together with the indices of the code vectors for the  
30 parameters in the system. This is suitable for large systems with a very large

number of users (ATM, access systems in companies). In this regard, see all Figure 7.

Another method is for the code vectors for the parameters for each speaker to be trained individually. The code vectors are then stored together with the values of the probability densities at those parameter space points defined by the code vectors. One possible way of carrying out this method is shown in Figure 6 (speaker identification using the parameters of an LPAS coder, probability densities are stored together with the code vectors for the parameters). This method is intended for a small number of speakers (for example, for a voice-controlled door in a dwelling).

#### Training phase of a speaker identification system

The probability density distributions for the speaker classes are estimated from the training material. For text-dependent speaker identification (speaker identification/speaker verification), a specific sentence or keyword is repeated during the training phase until the speaker identification operates reliably.

Phonetically balanced spoken material must be recorded for text-independent speaker verification. In this case as well, the training phase must be repeated until the speaker identification/verification operates reliably.

The material recorded during the training phase is in each case used with a phase shift a number of times for training, in order to make the speaker identification system independent of the initial phase of the recorded voices. The data used for training are referred to as the training statement  $TS_{sp_i}$ , with  $sp_i$  symbolizing the speaker.

#### Estimation of the probability densities

In order to describe the method according to the present invention for estimation of the probability densities of the parameters for the speaker classes, a number of necessary definitions first will be introduced ~~first of all~~.

The introduced abstraction of the coding process has the advantage that the estimation of the probability densities can be described in a simple manner without needing to go into the details of the highly complicated operations in the speech

coder. A detailed description of the parameter calculation process can be found in [4] and [5].

A speech coder operates in evaluation intervals. The operations described in that section via the LPAS coder are carried out in the speech coder for each signal frame, and supply the parameters of the speech signal for the respective frame.

Calculation of a non-quantized parameter vector  $p$  from the signal frame  $x$  is written as  $p = K_p(x)$  in an open-loop optimization procedure. The quantization of the parameter is referred to as  $\hat{p} = Q_p(p)$ . That region in the parameter space of the parameter  $p$  which is mapped onto the code vector  $\hat{p}$  in the coding process is referred to as  $S_{\hat{p}} = \{p : Q_p(p) = \hat{p}\}$ . The volume of this region is referred to as  $V(S_{\hat{p}})$ .

The set of possible code vectors for the parameter  $p$  is written as  $C_p = \{\hat{p}_i; i=1..N_p\}$ , where  $N_p$  is the number of code vectors. The set or regions which are linked to the code vectors is referred to as  $R_p = \{S_i; i=1..N_p\}$ . The association function for a region  $S_i$  is referred to as:

$$1_{S_i}(p) = \begin{cases} 1 & \text{for } p \in S_i \\ 0 & \text{for } p \notin S_i \end{cases}$$

The frequency of occurrence of a parameter in the training statement is calculated using

20

$$f_{S_i} = \frac{\text{Number of parameter values from the training statement } TS_{sp_i} \text{ which occur in the region } S_i}{\text{Number of parameter values from the training statement } TS_{sp_i}}$$

The estimated probability density distribution then becomes:

25

$$p(p | sp_i) = \sum_{k=1}^{N_p} 1_{S_k}(p) \frac{f_{S_k}}{V(S_k)}$$

*Estimation of the probabilities*

The probability functions (probability mass functions) are estimated for those parameters which are regarded as discrete probability variables, that is to say in particular the excitation from the codebook, which is optimized using a closed-loop procedure, and the fundamental period of the speech signal. These probability functions are defined as the frequencies of the given parameter codes in the training statement for the respective speaker.

#### *Storage of the probability distributions*

The speech parameters are not all calculated at the same time, but successively, in a speech coder. For example, the short-term predictor parameters are calculated first, and the remaining parameters for already known short-term predictor parameters are then optimized with regard to the synthesis or the prediction error. This allows effective storage of the probability distributions as conditional probabilities of the code vectors in a tree structure. This is possible ~~thanks~~ due to the following relationship:

$$p(\mathbf{p}_K, \mathbf{p}_L, \mathbf{p}_A | \mathbf{sp}_i) = p(\mathbf{p}_K | \mathbf{sp}_i) p(\mathbf{p}_L | \mathbf{sp}_i, \mathbf{p}_K) p(\mathbf{p}_A | \mathbf{sp}_i, \mathbf{p}_K, \mathbf{p}_L)$$

$P_K$  - Vector for a short-term parameter

$P_L$  - Vector for a long-term parameter

$P_A$  - Vector for an excitation parameter

A major simplification can be achieved if the speech parameters within a signal frame can be assumed to be statistically independent. The above formula then becomes:

$$p(\mathbf{p}_K, \mathbf{p}_L, \mathbf{p}_A | \mathbf{sp}_i) = p(\mathbf{p}_K | \mathbf{sp}_i) p(\mathbf{p}_L | \mathbf{sp}_i) p(\mathbf{p}_A | \mathbf{sp}_i)$$



The probability densities need to be stored at a very large number of points in parameter space in the system. The number of bits used for storing probability densities is critical to the complexity of the overall system. A vector quantizer is therefore used for the probability values. This makes it possible to reduce the number of bits used for storing the probability distributions.

*System safety and security*

In order to prevent the system from being outwitted, noise is transmitted at the same time that the voice of the speaker is being recorded, which noise is known to the system, and from which the digitized speech signal is subtracted.

The present invention can be used for access control applications, such as voice-controlled doors, or for verification; for example, for bank access systems. The procedure can be implemented as a program module on a processor which carries out the task of speaker identification in the system.

~~An exemplary embodiment of the invention is described with reference to Figures 7 and 8a to 8m.~~

Although the present invention has been described with reference to specific embodiments, those of skill in the art will recognize that changes may be made thereto without departing from the spirit and scope of the invention as set forth in the hereafter appended claims.

- [1] S. Furui, "Recent advances in speaker recognition", Pattern Recognition Letters, Tokyo Inst. of Technol., 1997
- [2] P. Vary, U. Heute, W. Hess, *Digitale Sprachsignalverarbeitung [Digital speech signal processing]*, B.G. Teubner, Stuttgart, 1998
- 5 [3] K. Kroschel, *Statistische Nachrichtentheorie [Statistical information theory]*, 3rd ed., Springer-Verlag, 1997
- [4] W.B. Kleijn, K.K. Paliwal, *Speech Coding and Synthesis*, Elsevier, 1995
- [5] ISO/IEC 14496-3, MPGA-3 HVXC Speech Coder description
- [6] Prakasa Rao, *Functional Estimation*, Academic Press, 1982

10

~~Abstract~~

**ABSTRACT OF THE DISCLOSURE**

~~Method for identification of speakers on the basis of their voices~~

- The invention relates to a A method for speaker identification using
- 5 parameters of an LPAS coder or of a parametric coder for modeling the probability distribution for the speaker classes.

1999P02665

Amended application version nat./reg, phase

Description

Method for identification of speakers on the basis of their voices

5

The invention relates to a method for identification of speakers on the basis of their voices.

10

The object on which the invention is based is to specify a method for identification of speakers on the basis of their voices, which method is robust, safe, secure and reliable.

15

According to the invention, this object is achieved by the features specified in patent claim 1.

20

The invention will be described in more detail in the following text using a flowchart.

1.

25

The invention allows the identification of the speaker on the basis of his voice. The problem of speaker identification is to distinguish between different speakers or to check the predetermined speaker identity, with the only input information being the recording of the voice of the speaker.

30

Furthermore, a method is proposed which prevents the access system from being outwitted when the voice and the keyword are recorded by third parties.

35

When complex probability distributions for the speech parameters of a speaker are stored, a compromise must be made between accuracy and memory requirement. For this reason, methods for storage of the probability distributions have been proposed which can be used as a function of the number of speakers.

2.

Until now, the speaker has been identified, for example, with the aid of hidden Markov models or by vector quantization, see reference [1].

5

3.

The invention solves the problem of speaker identification based on the parameters of an analysis by means of synthesis coders using linear prediction (LPAS) [1] (for example a harmonic vector excited codec [5] or waveform interpolation codec [4]). The speech signal parameters used in the past, such as Cepstral AR parameters, do not provide a satisfactory solution to the problem. For this reason, other parameters need to be used, such as parameters relating to the excitation of the vocal tract, which include information that is dependent on the speaker and is at the same time largely phoneme-independent.

Furthermore, the method for estimation of the probability distribution of the coder parameters for the respective speaker is given, as well as a method which prevents the access system from being outwitted.

#### 25 Speaker identification

In systems for *speaker identification*, statistical principles [2] are used to check whether the spoken sentence has been spoken by one of the speakers covered by the speaker identification system. In the process, there are in principle two types of speaker identification systems, text-dependent systems and text-independent systems. For the procedure described in the invention, text independence of the system is achieved by means of an expanded training phase, in which the speaker has to record a wide range of material, and the probability distributions of said

speech signal parameters are established from all the  
spoken material. A text-dependent system can be trained  
more easily since the spoken material which is spoken  
by the speaker during the usage phase is limited to a  
5 number of keywords or specific

sentences. The preparation phase is continued until the system reliably identifies the voice of the speaker.

The object of *speaker identification* is illustrated in Figure 1 (Problem of speaker identification).

*Speaker identification* is dealt with as a problem relating to the detection of multiples [2]. The classes to be distinguished between, one for each speaker who is intended to be identified by the system, are referred to as  $sp_i = 1..M$ , where  $M$  is the number of speakers covered by the speaker identification system. Speaker identification is based on recorded signals spoken by the respective speaker. The speech signal is segmented into the signal frames  $x = [x(1)..x(K)]$  ( $K = 160$  for a signal frame with a length of 20 ms and a sampling frequency of 8 kHz, for example). The segmentation process produces the speech signal frames  $x(1)..x(N)$ , where  $N$  depends on the total length of the sentence or keyword spoken by the speaker. The decision on the speaker is made from the probabilities or probability densities (referred to jointly as probability scores) that the vectors of the samples  $x(l)$   $l = 1..N$  belong to the class  $sp_i$ . The statistically optimum decision scheme selects that class  $sp_i$  having the highest probability value for given  $x(l)$ ,  $l = 1..N$ . This means that the vector  $x(l)$  is assigned to the class  $sp_j$  for which:

$$p(x(1)...x(N) | sp_j) > p(x(1)...x(N) | sp_i) \text{ for all } j \neq i$$

#### Speaker verification

The problem of *speaker verification* is to check the predetermined identity of the speaker on the basis of his voice. This corresponds to the situation

illustrated in Figure 2 (Problem of speaker verification).

The process of speaker verification is carried out in a similar manner to that of speaker identification, that



is to say the spoken sentence is likewise segmented. However, after this, the voice is not classified, but a probability score is calculated for the predetermined speaker identity, and is compared with a threshold. The  
5 identity of the speaker is thus confirmed on the basis of his voice when:

$$p(x(1)..x(N) | sp_j) > \text{threshold}$$

where  $sp_j$  corresponds to the predetermined speaker  
10 identity. The threshold must be set sufficiently high to avoid the situation in which a speaker with a different identity to that predetermined is accepted/authorized.

15 LPAS coder

The speech coding methods used nowadays are predominantly based on the analysis by synthesis method using an LPC synthesis filter [2]. In these methods, speech coding is optimized by repetition of the coding  
20 and decoding operations until the optimum parameter set is found for the given speech section.

One of the most widely used types of LPAS coder is the CELP coder. One relatively new development is the  
25 harmonic vector excited codec, where the form of the excitation signals is particularly suitable for the described task. Figure 3 (Design of an LPAS copier) illustrates the synthesis model of a CELP coder. The synthesis model defines the method for calculating the  
30 synthesized speech signal from the quantized parameters of the speech signal. In general, each LPAS coder has the following parameter groups:

- Short-term predictor parameters. The short-term  
35 predictor parameters are generally calculated by

- 4a -

means of classical LPC analysis, using the correlation method or the covariance method for linear prediction [3]. 8-10 LPC coefficients are used for signal frames with a length of 20 to 30 ms and a

AMENDED SHEET

sampling rate of 8 kHz. The short-term predictor parameters may occur in various forms (for example the reflection coefficients or in the form of line spectrum frequencies LSF), depending on the representation which can be quantized better. It has been found that the LSF coefficients are most suitable for quantization, and this form of prediction coefficients is generally used. The short-term predictor parameters are calculated using an open-loop procedure, that is to say without the overall optimization, illustrated in Figure 1, with the other parameters relating to the synthesis error.

- Long-term predictor parameters. Long-term predictor parameters are used in a filter which synthesizes the fundamental frequency of the speech signal. This is generally a long-term predictor with a filter coefficient and a parameter for the fundamental period of the voice signal. A long-term predictor with the parameters  $b = [b, N]$  is a part of Figure 2. The long-term predictor parameters are likewise calculated using an open-loop procedure, without overall optimization with the other parameters. In some coders, a refined search is sometimes carried out based on the long-term predictor parameters using a closed-loop procedure.

- The excitation parameters. The 5-10 ms subframes of the remaining signal are vector-quantized using a closed-loop procedure in a CELP coder. The transmitted parameters allow the signal forms to be reproduced at the decoder end from the stored codebook.

In an HVXC codec, the output from the LPC analysis filter is transformed to the frequency domain and the

- 5a -

fundamental-period-normalized spectral envelope is  
vector-quantized.

AMENDED SHEET

Speaker identification using the parameters of an LPAS coder

The speech coder parameters provide a comprehensive description of the possible speech signals using  
5 considerably fewer parameters than when the speech signal is represented as a sequence of samples.

The decomposition of the speech signal into the said parameter groups can be used in various ways for  
10 speaker identification. The methods for calculation of the parameters and synthesis of the speech signal imply probability density estimation methods (for example the probabilities of the parameters, which are regarded as discrete probability variables). Those defined using a  
15 closed-loop procedure should actually be regarded as discrete probability variables, since it is impossible to link the volumes of the parameter space regions of the vector quantizer for parameters such as these. This relates in particular to the excitation parameters. The  
20 probability distributions for such parameters are estimated by calculating relative frequencies of the parameters/code vectors in the training statement.

Those which are calculated using an open-loop procedure  
25 in the coder are initially available in a non-quantized form and are quantized only after this, with vector quantization generally being used. For parameters such as these, the probability densities can be estimated from the training statement. This approach is used  
30 primarily for the short-term predictor parameters.

The probability density estimation is based on the histogram method [6]. This method requires knowledge of the volumes of those regions of the parameter space  
35 which are linked to the quantized points.

- 6a -

A method for storage of probability distributions is obtained, according to Figure 5 (Speaker identification using

AMENDED SHEET

the parameters of an LPAS coder), if the possible code vectors for the speech signal parameters are stored once for the entire population, which corresponds to the situation where the quantization steps/code vectors are determined once, from the database which contains the recordings by a large number of speakers. The probability distributions of the parameters for the speakers are then stored together with the indices of the code vectors for the parameters in the system. This is suitable for large systems with a very large number of users (ATM, access systems in companies).

Another method is for the code vectors for the parameters for each speaker to be trained individually. The code vectors are then stored together with the values of the probability densities at those parameter space points defined by the code vectors. One possible way of carrying out this method is shown in Figure 6 (speaker identification using the parameters of an LPAS coder, probability densities are stored together with the code vectors for the parameters). This method is intended for a small number of speakers (for example for a voice-controlled door in a dwelling).

#### 25 Training phase of a speaker identification system

The probability density distributions for the speaker classes are estimated from the training material. For text-dependent speaker identification (speaker identification/speaker verification), a specific sentence or keyword is repeated during the training phase until the speaker identification operates reliably.

Phonetically balanced spoken material must be recorded for text-independent speaker verification. In this case as well, the training phase must be repeated until the speaker identification/verification operates reliably.

The material recorded during the training phase is in each case used with a phase shift a number of times for training, in order to make the speaker identification system independent of the initial phase of the recorded  
5 voices. The data used for training are referred to as the training statement  $TS_{sp_i}$ , with  $sp_i$  symbolizing the speaker.

*Estimation of the probability densities*

10 In order to describe the method according to the invention for estimation of the probability densities of the parameters for the speaker classes, a number of necessary definitions will be introduced first of all. The introduced abstraction of the coding process has  
15 the advantage that the estimation of the probability densities can be described in a simple manner without needing to go into details of the highly complicated operations in the speech coder. A detailed description of the parameter calculation process can be found in  
20 [4] and [5].

A speech coder operates in evaluation intervals. The operations described in that section via the LPAS coder are carried out in the speech coder for each signal  
25 frame, and supply the parameters of the speech signal for the respective frame.

Calculation of a non-quantized parameter vector  $p$  from the signal frame  $x$  is written as  $p = K_p(x)$  in an open-  
30 loop optimization procedure. The quantization of the parameter is referred to as  $\hat{p} = Q_p(p)$ . That region in the parameter space of the parameter  $p$  which is mapped onto the code vector  $\hat{p}$  in the coding process is referred to as  $S_p = \{p : Q_p(p) = \hat{p}\}$ . The volume of this region is  
35 referred to as  $V(S_p)$ .

The set of possible code vectors for the parameter  $p$  is written as  $C_p = \{\hat{p}_i; i=1..N_p\}$ , where  $N_p$  is the number of



code vectors. The set of regions which are linked to the code vectors is referred to as  $R_p = \{S_i; i=1..N_p\}$ . The association function for a region  $S_i$  is referred to as:

$$1_{S_i}(p) = \begin{cases} 1 & \text{for } p \in S_i \\ 0 & \text{for } p \notin S_i \end{cases}$$

The frequency of occurrence of a parameter in the training statement is calculated using

5

$$f_{S_i} = \frac{\text{Number of parameter values from the training statement } TS_{sp_i} \text{ which occur in the region } S_i}{\text{Number of parameter values from the training statement } TS_{sp_i}}$$

The estimated probability density distribution then becomes:

10

$$p(p | sp_i) = \sum_{k=1}^{N_p} 1_{S_k}(p) \frac{f_{S_k}}{V(S_i)}$$

#### *Estimation of the probabilities*

15

The probability functions (probability mass functions) are estimated for those parameters which are regarded as discrete probability variables, that is to say in particular the excitation from the codebook, which is optimized using a closed-loop procedure, and the fundamental period of the speech signal. These probability functions are defined as the frequencies of the given parameter codes in the training statement for the respective speaker.

20

#### *Storage of the probability distributions*

25

The speech parameters are not all calculated at the same time, but successively, in a speech coder. For example, the short-term predictor parameters are calculated first, and the remaining parameters for already known short-term predictor parameters are then optimized with regard to the synthesis or the prediction error. This allows effective storage of the probability distributions as conditional probabilities

30

- 9a -

of the code vectors in a tree structure. This is possible thanks to the following relationship:

AMENDED SHEET

$$p(p_K, p_L, p_A | sp_i) = p(p_K | sp_i) p(p_L | sp_i, p_K) p(p_A | sp_i, p_K, p_L)$$

- $p_K$  - Vector for a short-term parameter  
 $p_L$  - Vector for a long-term parameter  
5  $p_A$  - Vector for an excitation parameter

A major simplification can be achieved if the speech parameters within a signal frame can be assumed to be statistically independent. The above formula then  
10 becomes:

$$p(p_K, p_L, p_A | sp_i) = p(p_K | sp_i) p(p_L | sp_i) p(p_A | sp)$$

The probability densities need to be stored at a very  
15 large number of points in parameter space in the system. The number of bits used for storing probability densities is critical to the complexity of the overall system. A vector quantizer is therefore used for the probability values. This makes it possible to reduce  
20 the number of bits used for storing the probability distributions.

#### *System safety and security*

In order to prevent the system from being outwitted,  
25 noise is transmitted at the same time that the voice of the speaker is being recorded, which noise is known to the system, and from which the digitized speech signal is subtracted.

30 5.

The invention can be used for access control applications, such as voice-controlled doors, or for verification, for example for bank access systems. The procedure can be implemented as a program module on a

processor which carries out the task of speaker identification in the system.

An exemplary embodiment of the invention is described  
5 with reference to Figures 7 and 8a to 8m.

- [1] S. Furui, "Recent advances in speaker recognition", Pattern Recognition Letters, Tokyo Inst. of Technol., 1997
- 5 [2] P. Vary, U. Heute, W. Hess, *Digitale Sprachsignalverarbeitung [Digital speech signal processing]*, B.G. Teubner, Stuttgart, 1998
- [3] K. Kroschel, *Statistische Nachrichtentheorie [Statistical information theory]*, 3rd ed., Springer-Verlag, 1997
- 10 [4] W.B. Kleijn, K.K. Paliwal, *Speech Coding and Synthesis*, Elsevier, 1995
- [5] ISO/IEC 14496-3, MPGA-3 HVXC Speech Coder description
- 15 [6] Prakasa Rao, *Functional Estimation*, Academic Press, 1982

## Patent Claims

1. A method for identification of speakers on the basis of their voices, having the following features:
- 5 (a) in a preparation phase,
- (a1) k text-dependent or text-independent reference spoken expressions, which form a speaker-related training statement, from M
- 10 speakers are segmented into first speech signal frames of length L,
- (a2) the first speech signal frames are supplied to an analysis-by-synthesis coder based on linear prediction,
- 15 (a3) a first short-term predictor parameter, long-term predictor parameter and/or excitation parameter for the coder are/is calculated in the analysis-by-synthesis coder for each of the M
- speakers and for each first speech signal frame in each case, with the parameters then forming
- 20 speaker-related training material,
- (a4) the frequency of the respective occurrence of the first short-term predictor parameter, of the long-term predictor parameter and/or of the
- 25 excitation parameter for the coder in the speaker-related training statement and/or the probability densities with which the first short-term predictor parameter, the long-term predictor parameter and/or the excitation parameter are/is
- 30 contained in the speaker-related training statement are/is calculated in the analysis-by-synthesis coder for each of the M speakers and for each first speech signal frame in each case,
- (a5) the calculated frequencies and/or probability
- 35 densities are stored on a speaker-related basis as speaker data,

(b) in a simulated usage phase of the training phase,

(b1) a text-dependent or text-independent simulation spoken expression of an m-th speaker where  $m=1..M$  is segmented into second speech signal frames of length L,

(b2) the second speech signal frames are supplied to the analysis-by-synthesis coder,

(b3) a second short-term predictor parameter, long-term predictor parameter and/or excitation parameter for the coder are/is calculated

in the analysis-by-synthesis coder for the m-th speaker and for every other speech signal frame in each case,



(b4) first probability hits are calculated for every other speech signal frame from the calculated second short-term predictor parameter, long-term predictor parameter and/or excitation parameter and the speaker data stored for the m-th speaker in the preparation phase, which probability hits indicate the probability with which the second short-term predictor parameter, long-term predictor parameter and/or excitation parameter match(es) the first short-term predictor parameter, long-term predictor parameter and/or excitation parameter,

(b5) the first probability scores from all the second speech signal frames are combined,

(b6) a check is carried out to determine whether the combined first probability scores are greater than a predetermined first threshold which confirms the voice of the m-th speaker, when the combined first probability scores are greater than the predetermined first threshold or the preparation phase continues for a further i reference spoken expressions by the m-th speaker until the voice of the m-th speaker is confirmed, when the combined first probability scores are less than or equal to, or are less than, the predetermined first threshold,

(c) in a usage phase

(c1) a text-dependent or text-independent used spoken expression of the m-th speaker where  $m=1..M$  is segmented into third speech signal frames of length L,

(c2) the third speech signal frames are supplied to the analysis-by-synthesis coder,

(c3) a third short-term predictor parameter, long-term predictor parameter and/or excitation parameter for the coder are/is calculated in the analysis-by-synthesis coder for the m-th speaker

and for every third speech signal frame in each case,

- 5 (c4) second probability hits are calculated for every third speech signal frame from the calculated third short-term predictor parameter, long-term predictor parameter and/or excitation parameter and the speaker data stored for the

AMENDED SHEET

m-th speaker in the preparation phase, which second probability hits indicate the probability with which the third short-term predictor parameter, long-term predictor parameter and/or excitation parameter have been spoken by the m-th speaker,

(c5) the second probability hits from all the third speech signal frames are combined,

(c6) a check is carried out to determine whether the combined second probability scores are greater than a predetermined second threshold and the voice of the m-th speaker is identified when the combined second probability hits are greater than the predetermined second threshold, or the voice of the m-th speaker is not identified when the combined second probability scores are less than or equal to, or are less than, the predetermined second threshold.

2. The method as claimed in claim 1, characterized in that  
a harmonic vector excited predictive coder or a waveform interpolating coder is used, in particular, as a parametric coder.
3. The method as claimed in claim 1, characterized in that  
a coder based on linear prediction, in particular an LPAS coder, is used as the analysis-by-synthesis coder.
4. The method as claimed in one of claims 1 to 3, characterized in that  
the frequencies and/or probability densities are quantized using a vector quantizer having a specific, considerably reduced, number of bits.

5. The method as claimed in one of claims 1 to 4,  
characterized in that

noise which is known to the speaker identification system is also entered when the spoken expression of the speaker is entered into the speaker identification system.

5

6. The method as claimed in one of claims 1 to 5, characterized in that the noise which is also entered is subtracted internally, before the segmentation, from the recording of the speaker voice.

10

1999P02665

Abstract

Method for identification of speakers on the basis of their voices

The invention relates to a method for speaker identification using parameters of an LPAS coder or of a parametric coder for modeling the probability distribution for the speaker classes.

AMENDED SHEET

BOX PCT  
IN THE UNITED STATES ELECTED/DESIGNATED OFFICE  
OF THE UNITED STATES PATENT AND TRADEMARK OFFICE  
UNDER THE PATENT COOPERATION TREATY-CHAPTER II

SUBMISSION OF DRAWINGS

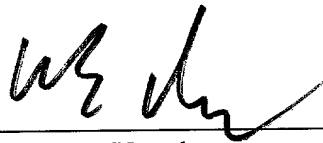
APPLICANTS: Marcin Kuropatwinski DOCKET NO: 112740-213  
SERIAL NO: 09/830,497 GROUP ART UNIT:  
EXAMINER:  
INTERNATIONAL APPLICATION NO: PCT/DE00/02917  
INTERNATIONAL FILING DATE: 25 August 2000  
INVENTION: METHOD FOR THE IDENTIFICATION OF SPEAKERS ON  
THE BASIS OF THEIR VOICES

Assistant Commissioner for Patents,  
Washington, D.C. 20231

Sir:

Applicant herewith submits nineteen sheets (Figs. 1-8m) of drawings for the  
above-referenced PCT application.

Respectfully submitted,

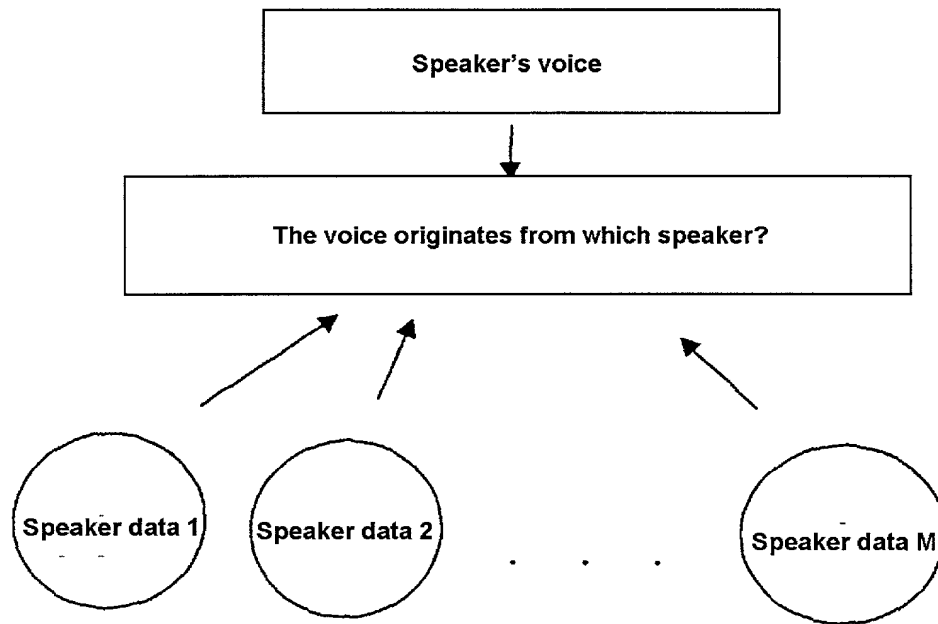


(Reg. No. 39,056)

William E. Vaughan  
Bell, Boyd & Lloyd LLC  
P.O. Box 1135  
Chicago, Illinois 60690-1135  
(312) 807-4292  
Attorneys for Applicant

1/19

FIG 1





2/19

FIG 2

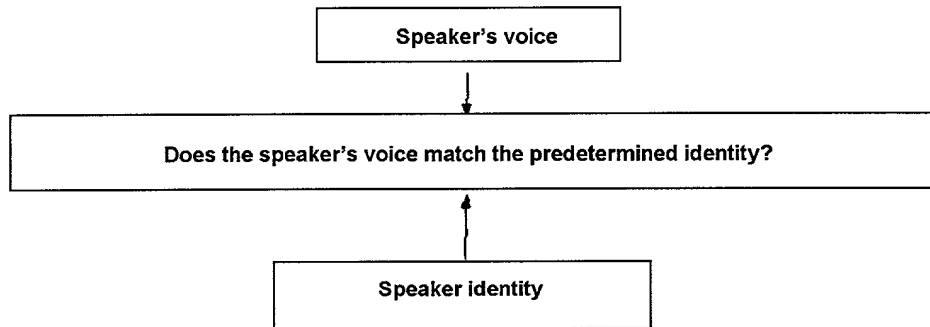
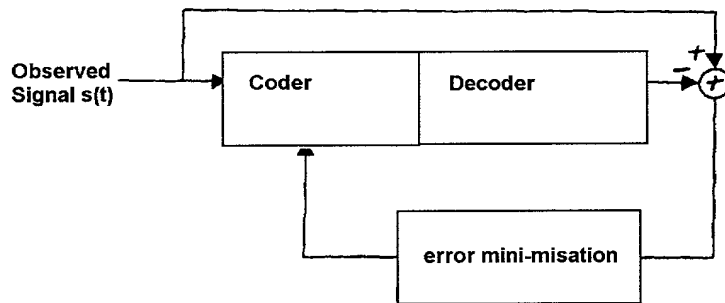
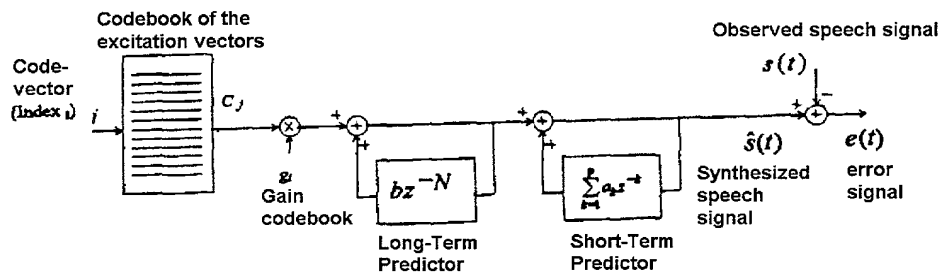


FIG 3



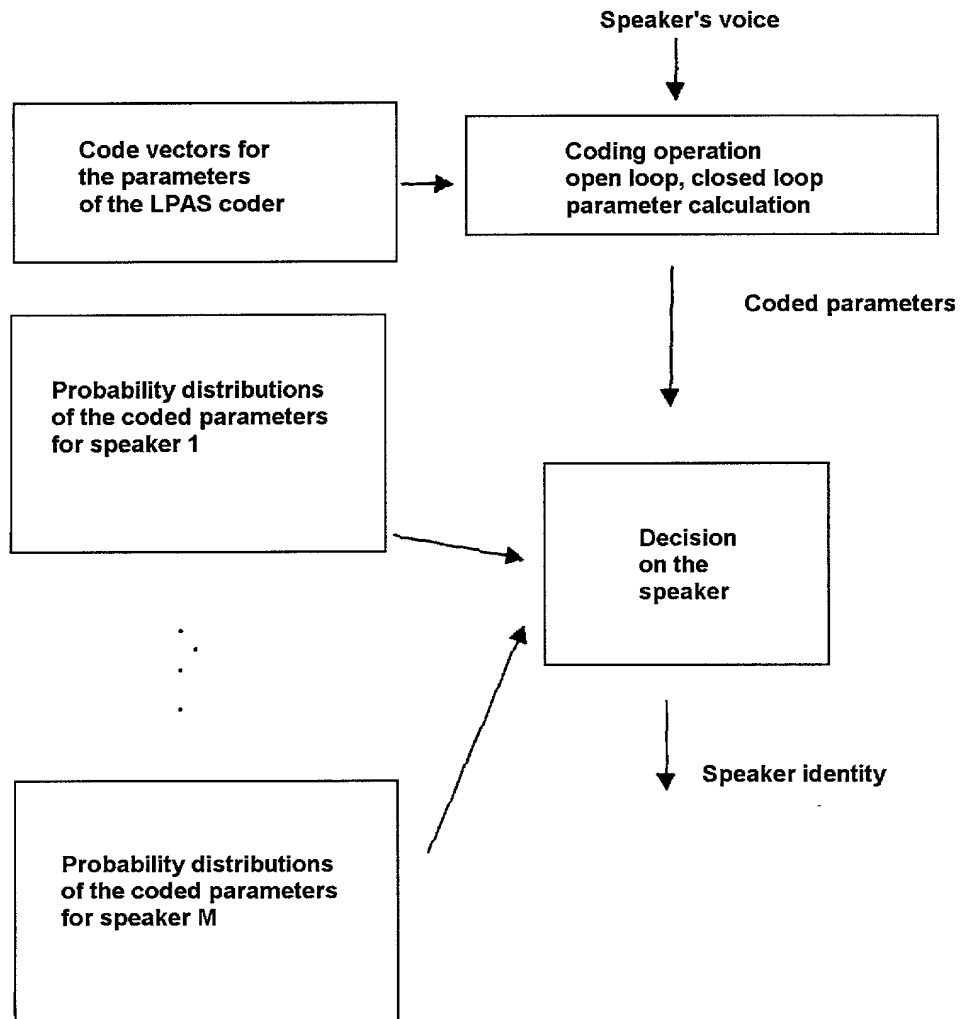
3/19

FIG 4



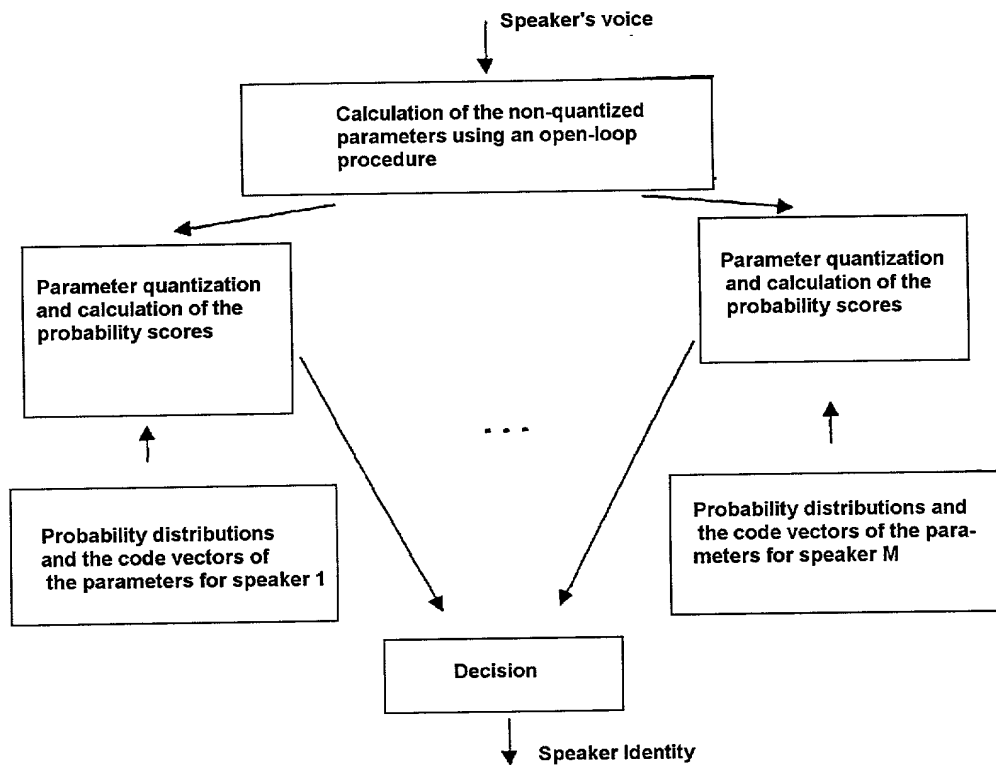
4/19

FIG 5



5/19

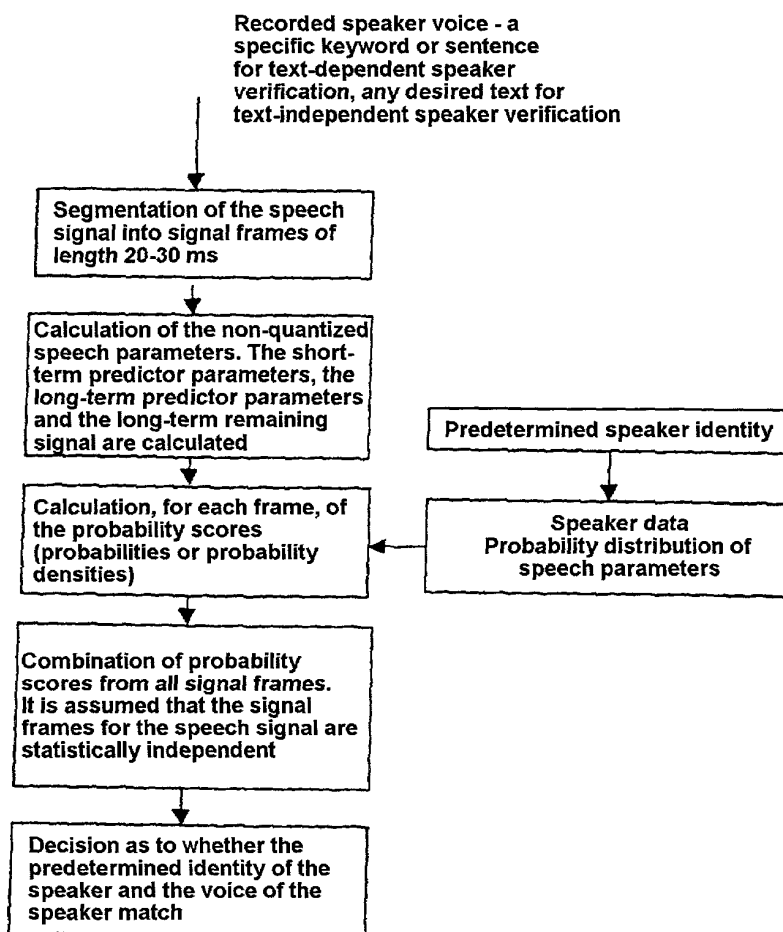
FIG 6



6/19

FIG 7

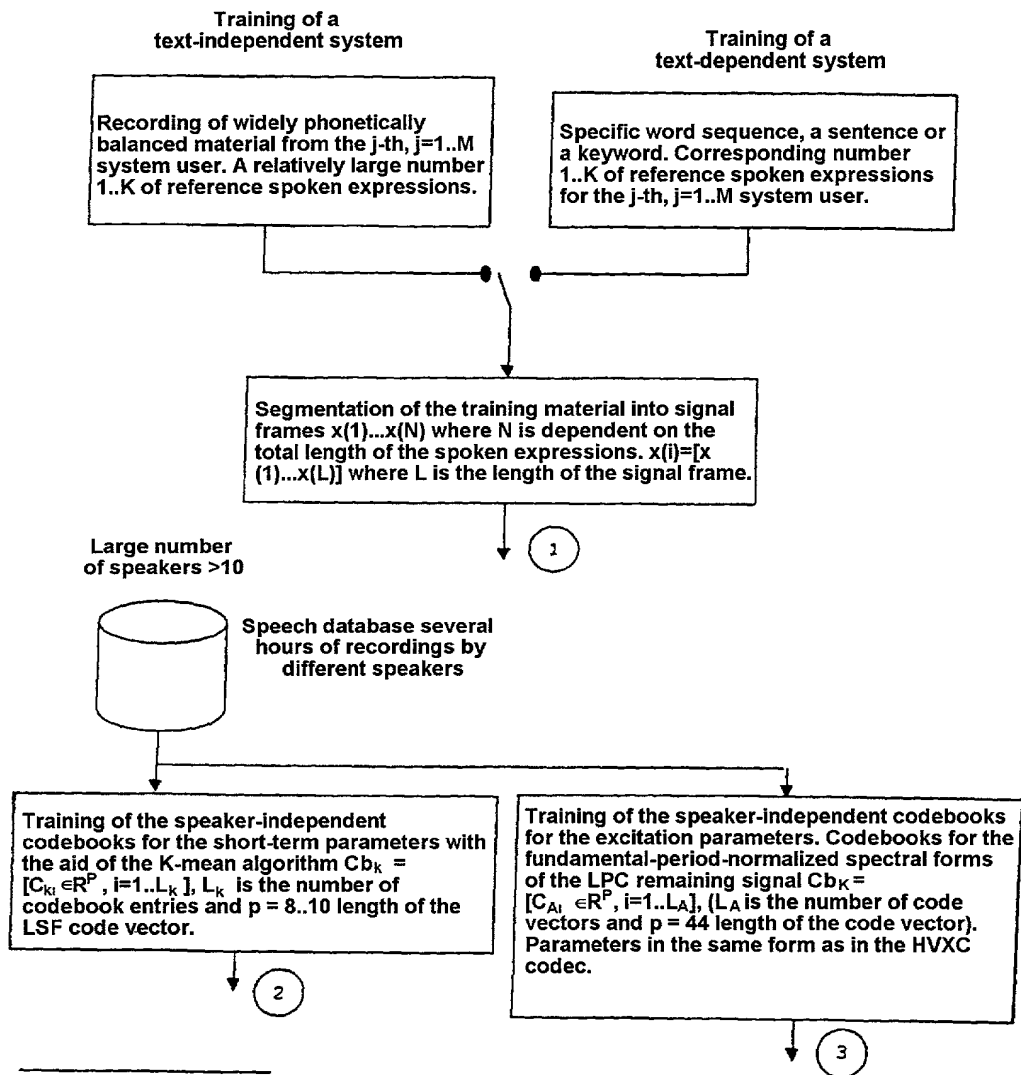
## Speaker verification using the parameters of an LPAS coder



7/19

FIG 8a

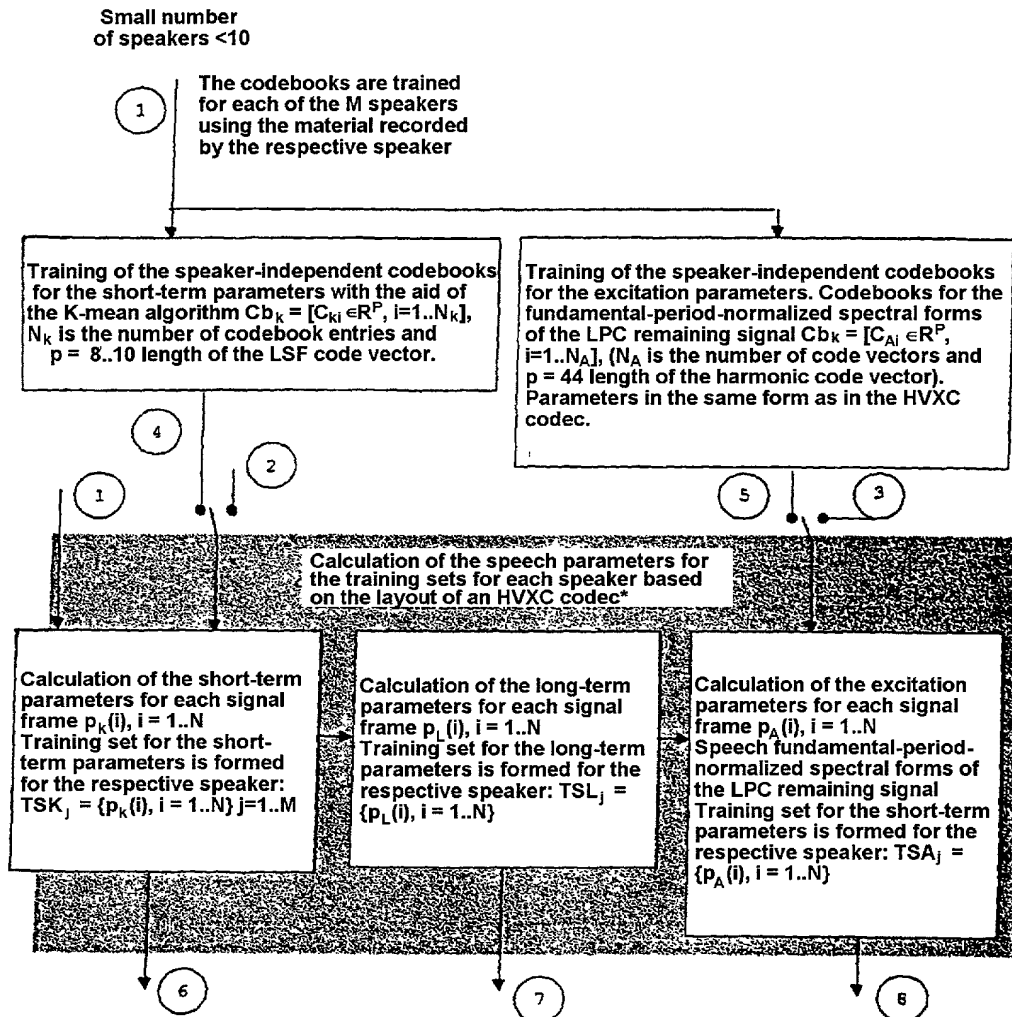
**Speaker identification system preparation phase\***  
(Profile for speaker j)



\* The process defined from hereon is carried out for each new user of the speaker identification system. The aim of the preparation phase is to produce the speaker data for each of the  $M$  speakers.

8/19

FIG 8b

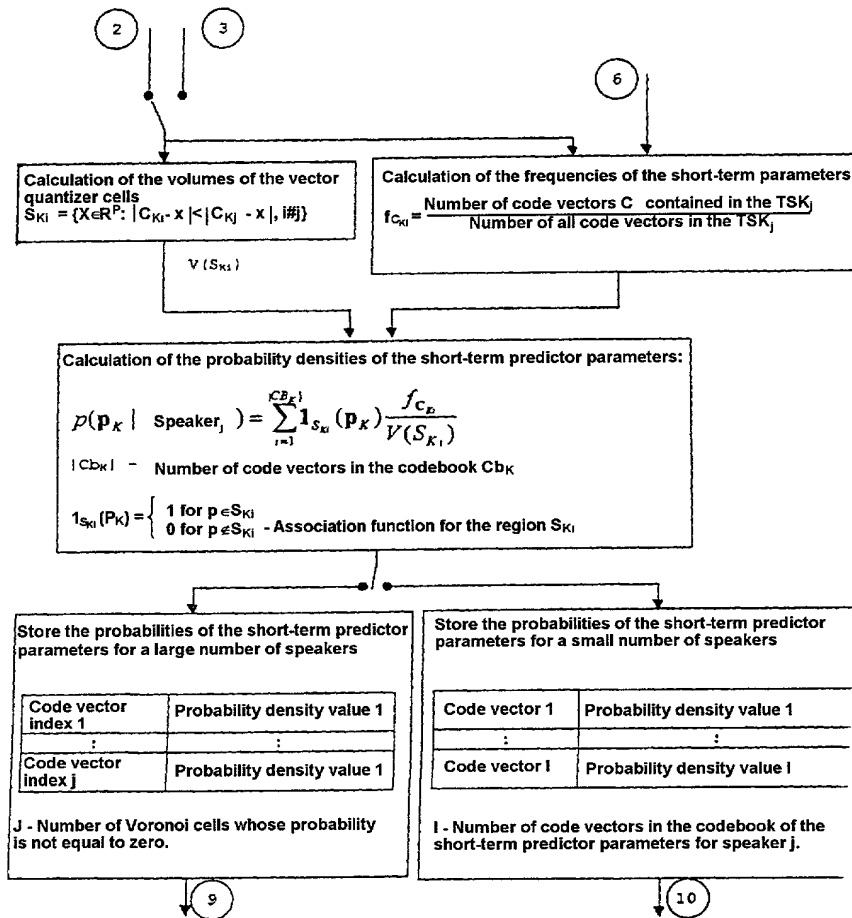


\* ISO/IEC 14496-3 Information Technology - Very low bit rate audio-visual coding

9/19

FIG 8c

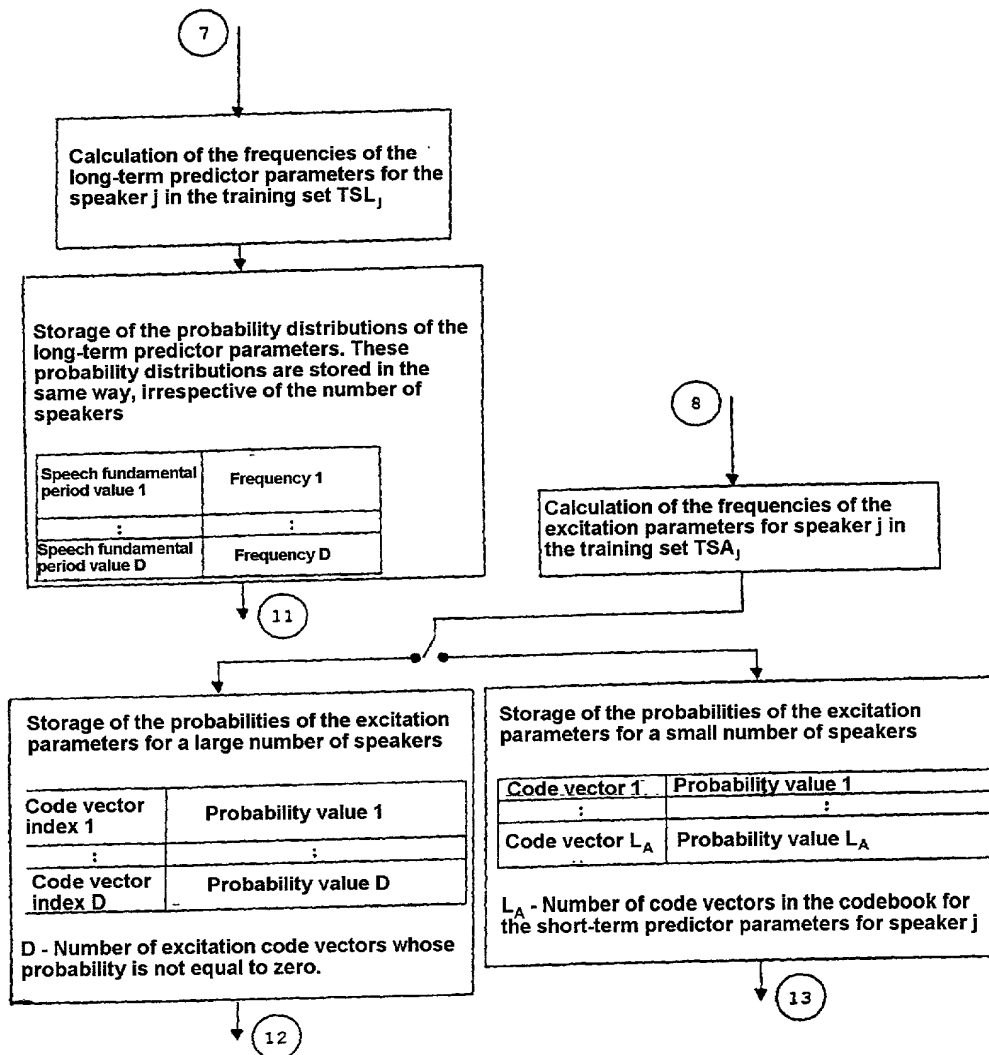
Calculation of the volumes of the Voronoi cell regions for the probability density estimate of the short-term predictor parameters





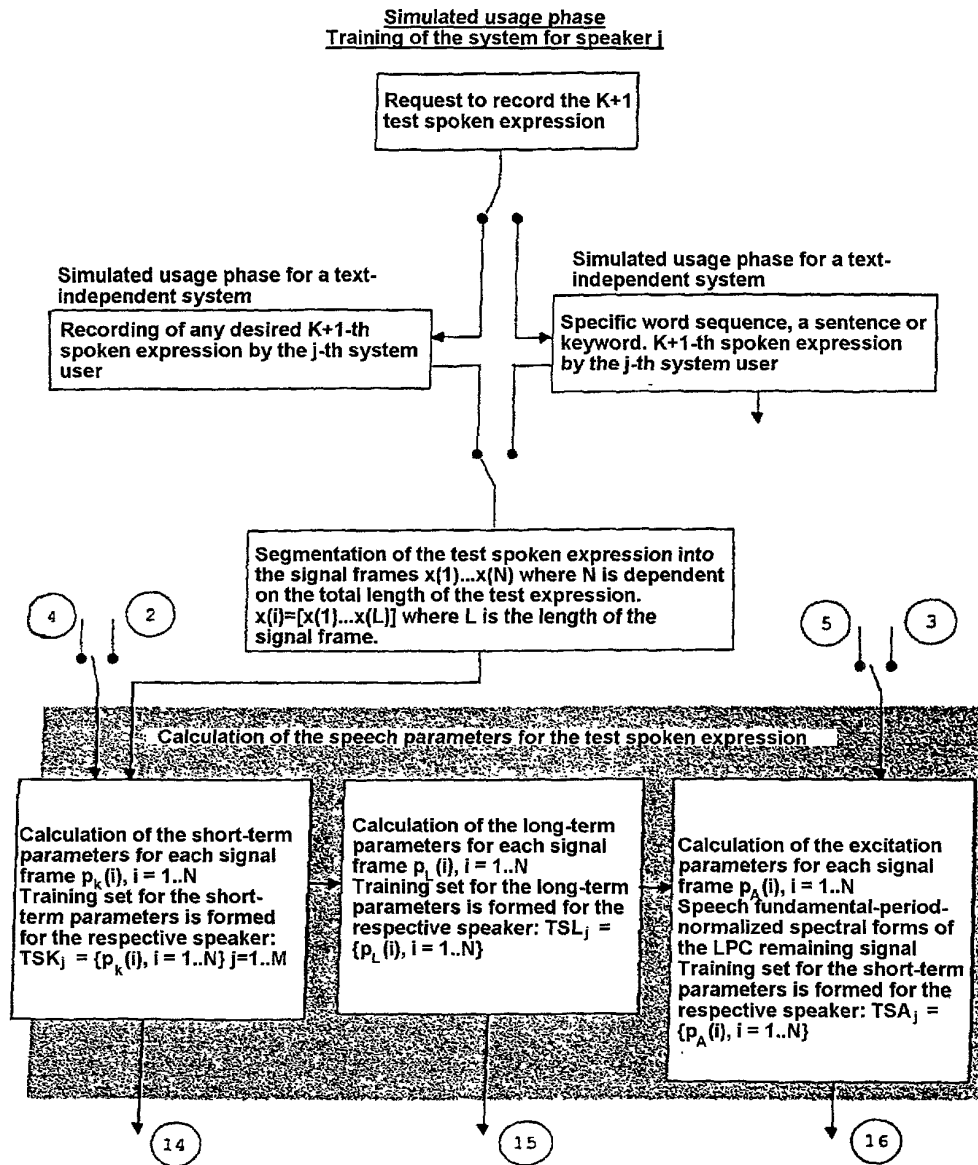
10/19

FIG 8d



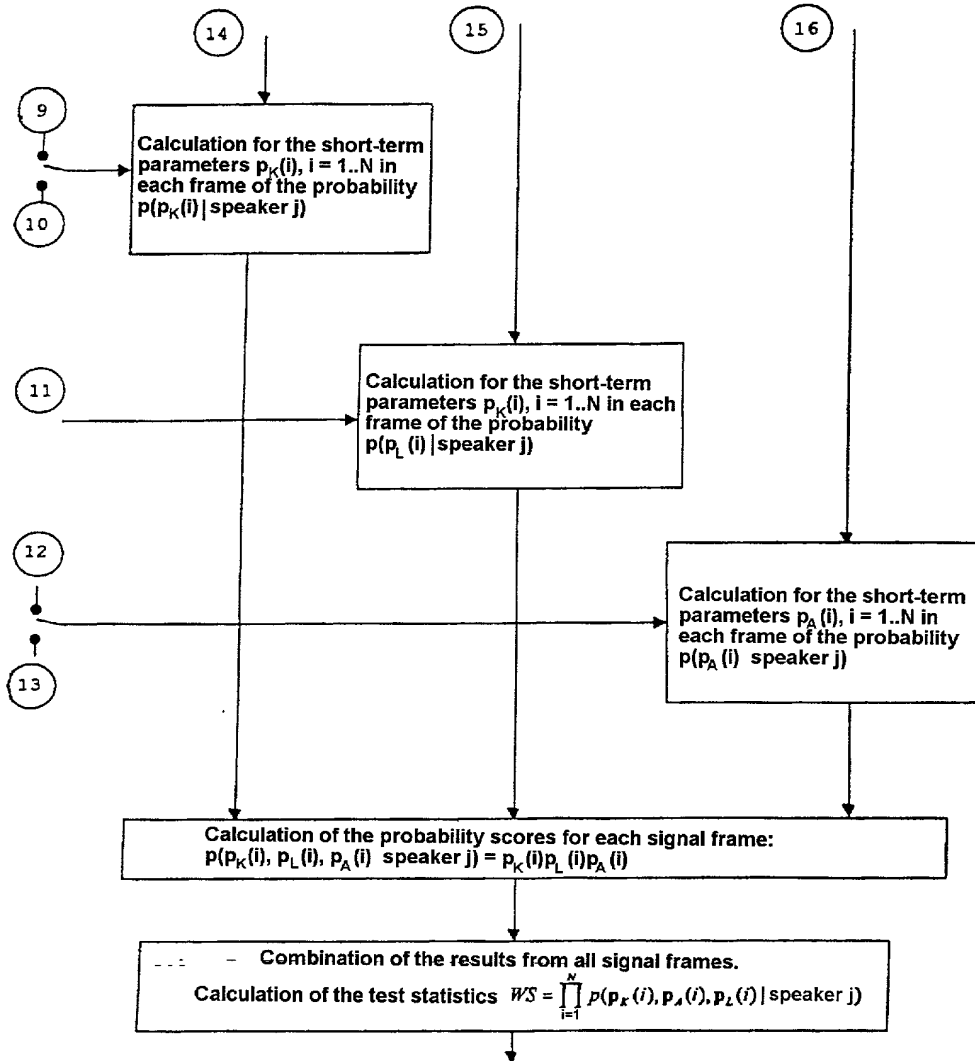
11/19

FIG 8e



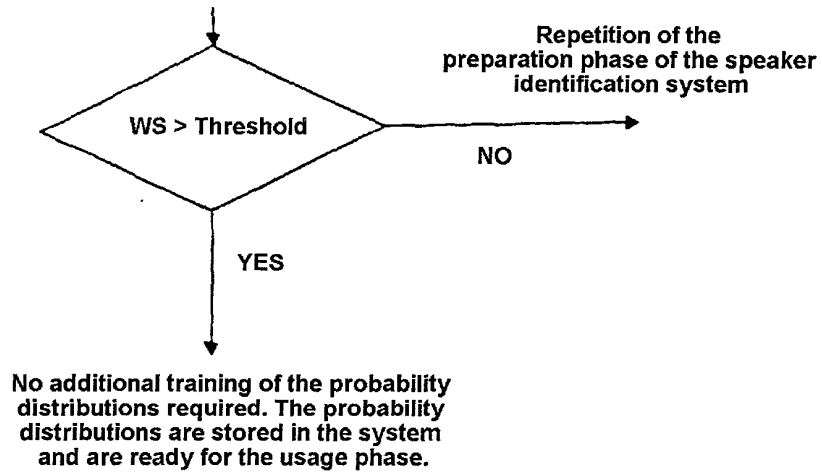
12/19

FIG 8f



13/19

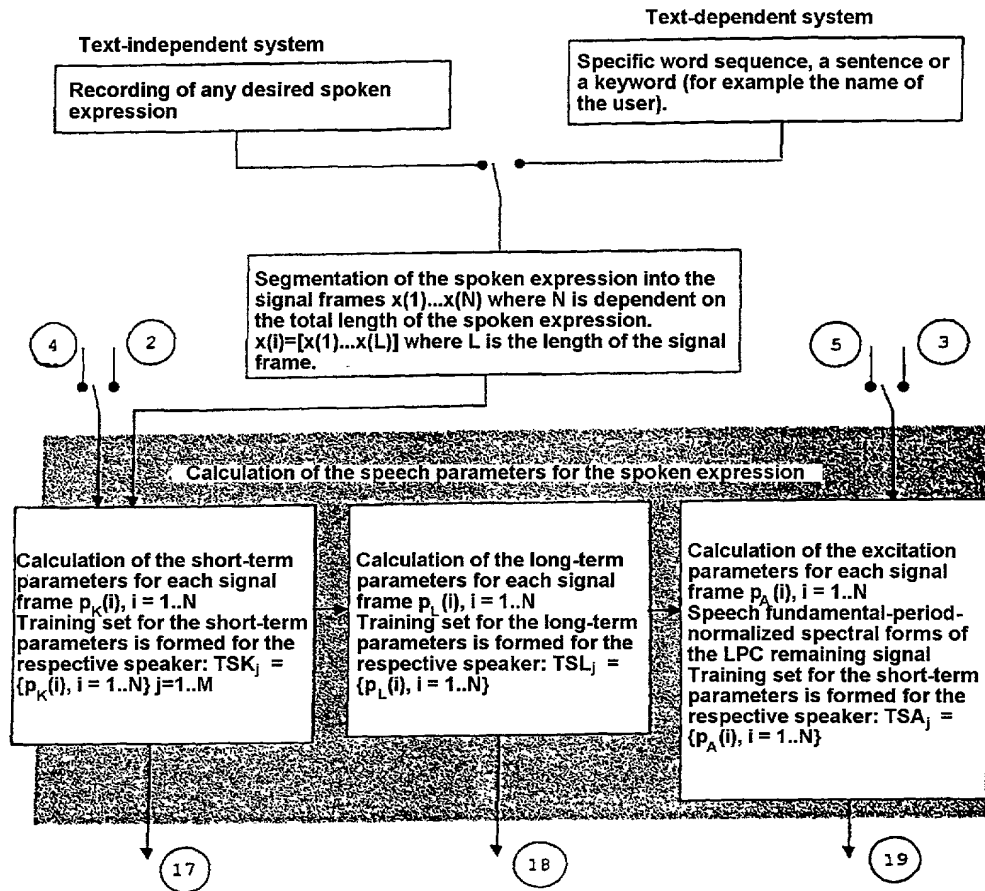
FIG 8g



14/19

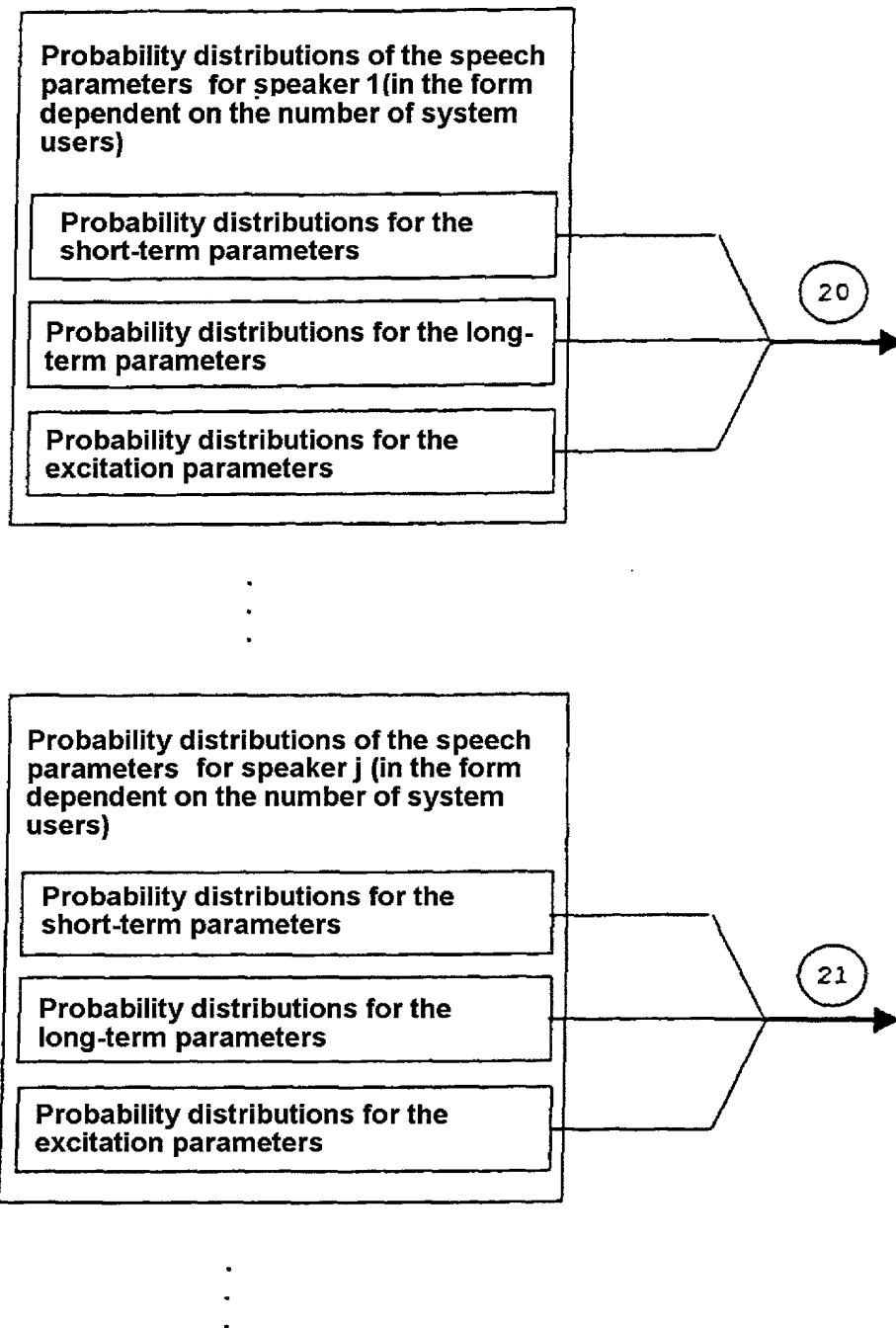
FIG 8h

**Speaker identification system usage phase**  
(Profile for the speaker j)



15/19

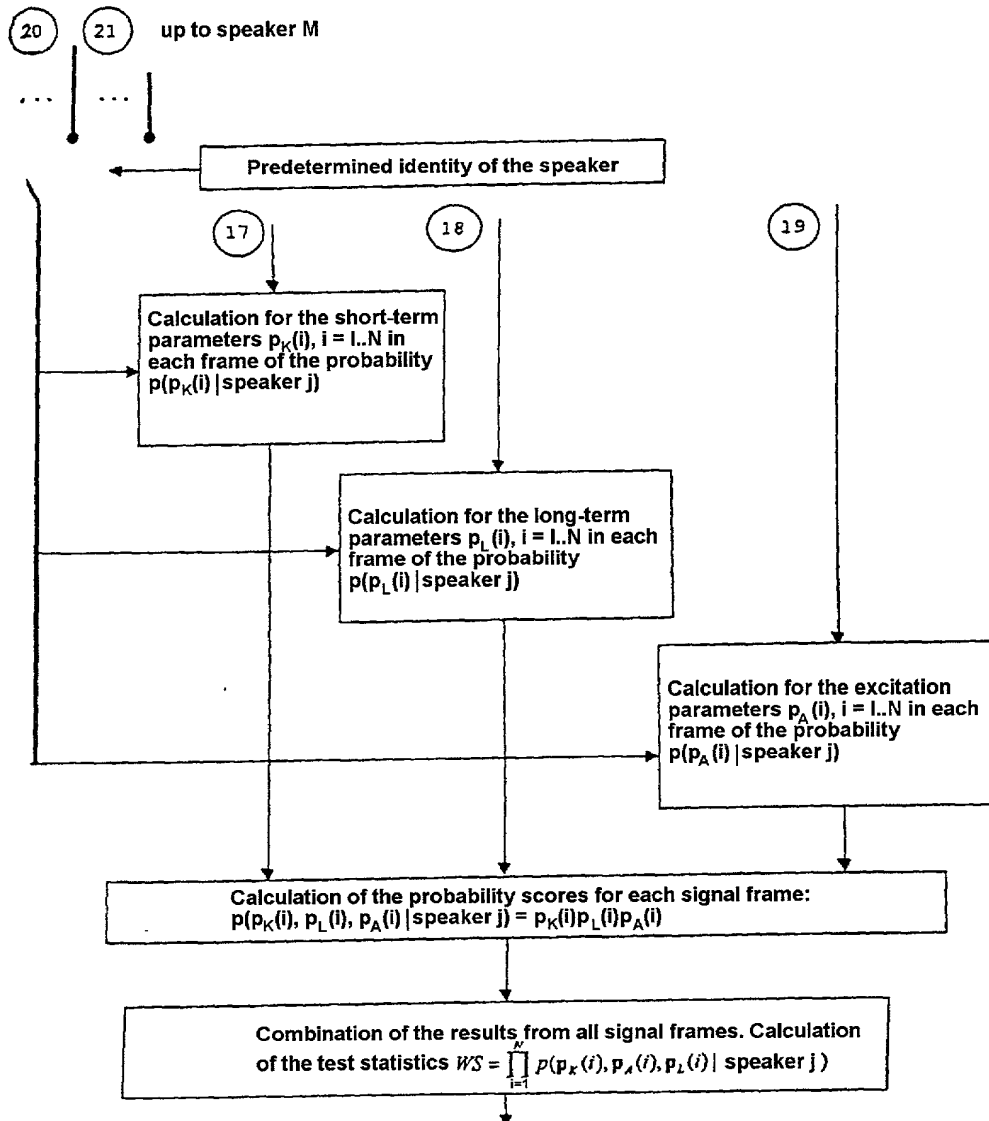
FIG 8i



up to speaker M

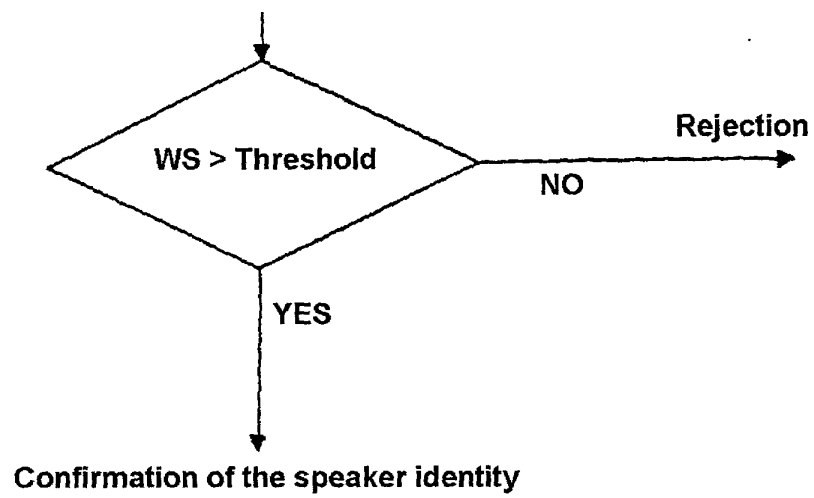
16/19

FIG 8j

Speaker verification

17/19

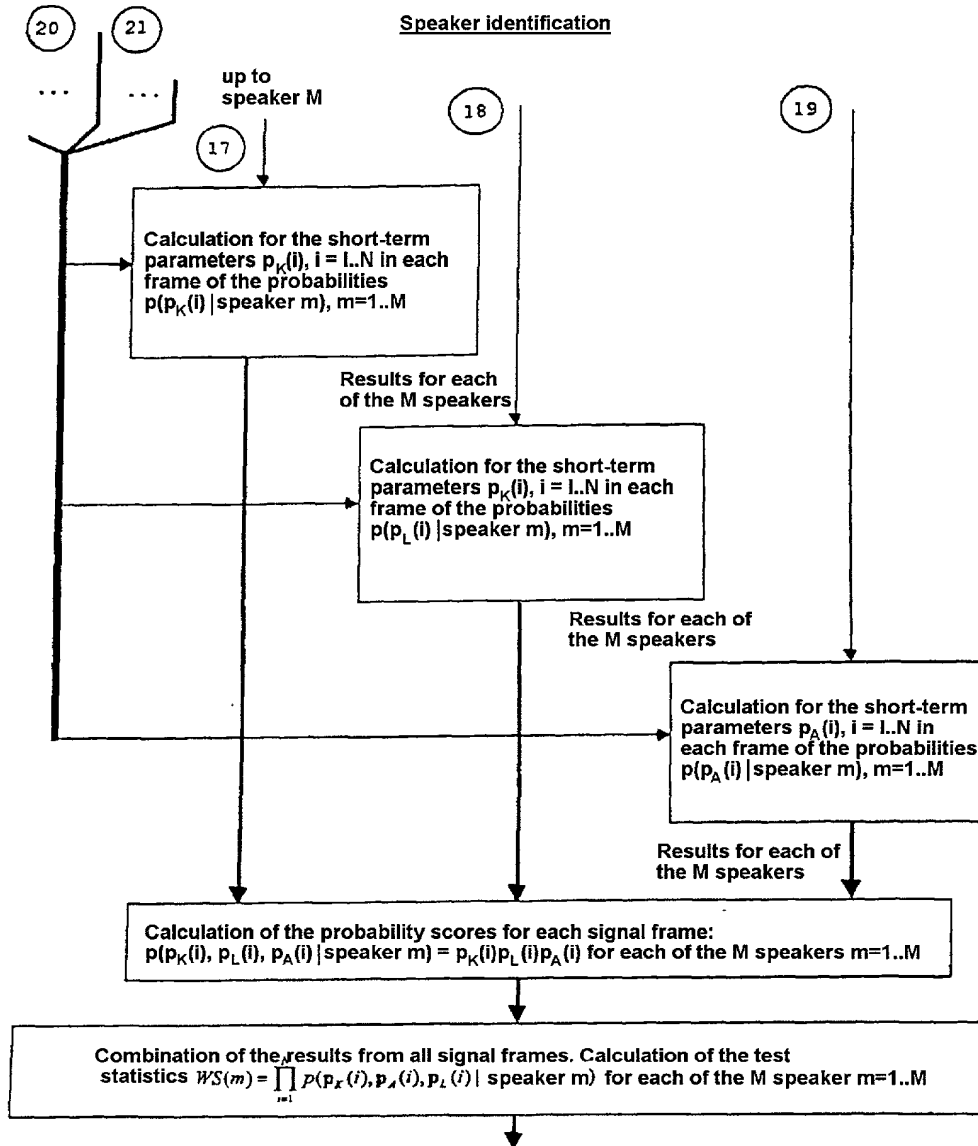
FIG 8k





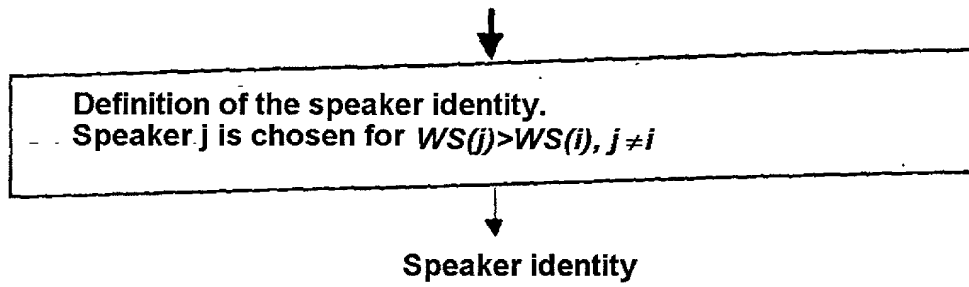
18/19

FIG 81



19/19

FIG 8m



# Declaration and Power of Attorney For Patent Application

## Erklärung Für Patentanmeldungen Mit Vollmacht

### German Language Declaration

Als nachstehend benannter Erfinder erkläre ich hiermit an Eides Statt:

As a below named inventor, I hereby declare that:

dass mein Wohnsitz, meine Postanschrift, und meine Staatsangehörigkeit den im Nachstehenden nach meinem Namen aufgeführten Angaben entsprechen,

My residence, post office address and citizenship are as stated below next to my name,

dass ich, nach bestem Wissen der ursprüngliche, erste und alleinige Erfinder (falls nachstehend nur ein Name angegeben ist) oder ein ursprünglicher, erster und Miterfinder (falls nachstehend mehrere Namen aufgeführt sind) des Gegenstandes bin, für den dieser Antrag gestellt wird und für den ein Patent beantragt wird für die Erfindung mit dem Titel:

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

Verfahren zum Trainieren eines  
Sprechererkennungssystems

Method for training a speaker recognition  
system

deren Beschreibung

the specification of which

(zutreffendes ankreuzen)

☐ hier beigefügt ist.

☒ am 25.08.2000 als

PCT internationale Anmeldung

PCT Anmeldeungsnummer PCT/DE00/02917

eingereicht wurde und am \_\_\_\_\_

abgeändert wurde (falls tatsächlich abgeändert).

(check one)

☐ is attached hereto.

☒ was filed on 25.08.2000 as

PCT international application

PCT Application No. PCT/DE00/02917

and was amended on \_\_\_\_\_  
(if applicable)

Ich bestätige hiermit, dass ich den Inhalt der obigen Patentanmeldung einschliesslich der Ansprüche durchgesehen und verstanden habe, die eventuell durch einen Zusatzantrag wie oben erwähnt abgeändert wurde.

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims as amended by any amendment referred to above.

Ich erkenne meine Pflicht zur Offenbarung irgendwelcher Informationen, die für die Prüfung der vorliegenden Anmeldung in Einklang mit Absatz 37, Bundesgesetzbuch, Paragraph 1.56(a) von Wichtigkeit sind, an.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

Ich beanspruche hiermit ausländische Prioritätsvorteile gemäss Abschnitt 35 der Zivilprozessordnung der Vereinigten Staaten, Paragraph 119 aller unten angegebenen Auslandsanmeldungen für ein Patent oder eine Erfindersurkunde, und habe auch alle Auslandsanmeldungen für ein Patent oder eine Erfindersurkunde nachstehend gekennzeichnet, die ein Anmeldedatum haben, das vor dem Anmeldedatum der Anmeldung liegt, für die Priorität beansprucht wird.

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

## German Language Declaration

Prior foreign applications  
Priorität beansprucht

Priority Claimed

19940567.0

DE

26.08.1999

☒

☐

(Number)  
(Nummer)

(Country)  
(Land)

(Day Month Year Filed)  
(Tag Monat Jahr eingereicht)

Yes  
Ja

No  
Nein

(Number)  
(Nummer)

(Country)  
(Land)

(Day Month Year Filed)  
(Tag Monat Jahr eingereicht)

☐  
Yes  
Ja

☐  
No  
Nein

(Number)  
(Nummer)

(Country)  
(Land)

(Day Month Year Filed)  
(Tag Monat Jahr eingereicht)

☐  
Yes  
Ja

☐  
No  
Nein

Ich beanspruche hiermit gemäss Absatz 35 der Zivilprozessordnung der Vereinigten Staaten, Paragraph 120, den Vorzug aller unten aufgeführten Anmeldungen und falls der Gegenstand aus jedem Anspruch dieser Anmeldung nicht in einer früheren amerikanischen Patentanmeldung laut dem ersten Paragraphen des Absatzes 35 der Zivilprozessordnung der Vereinigten Staaten, Paragraph 122 offenbart ist, erkenne ich gemäss Absatz 37, Bundesgesetzbuch, Paragraph 1.56(a) meine Pflicht zur Offenbarung von Informationen an, die zwischen dem Anmeldedatum der früheren Anmeldung und dem nationalen oder PCT internationalen Anmeldedatum dieser Anmeldung bekannt geworden sind.

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §122, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application.

PCT/DE00/02917

(Application Serial No.)  
(Anmeldeseriennummer)

25.08.2000

(Filing Date D, M, Y)  
(Anmeldedatum T, M, J)

(Status)  
(patentiert, anhängig,  
aufgegeben)

(Status)  
(patented, pending,  
abandoned)

(Application Serial No.)  
(Anmeldeseriennummer)

(Filing Date D,M,Y)  
(Anmeldedatum T, M; J)

(Status)  
(patentiert, anhängig,  
aufgeben)

(Status)  
(patented, pending,  
abandoned)

Ich erkläre hiermit, dass alle von mir in der vorliegenden Erklärung gemachten Angaben nach meinem besten Wissen und Gewissen der vollen Wahrheit entsprechen, und dass ich diese eidesstattliche Erklärung in Kenntnis dessen abgebe, dass wissentlich und vorsätzlich falsche Angaben gemäss Paragraph 1001, Absatz 18 der Zivilprozessordnung der Vereinigten Staaten von Amerika mit Geldstrafe belegt und/oder Gefängnis bestraft werden koennen, und dass derartig wissentlich und vorsätzlich falsche Angaben die Gültigkeit der vorliegenden Patentanmeldung oder eines darauf erteilten Patentes gefährden können.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

## German Language Declaration

**VERTRETUNGSVOLLMACHT:** Als benannter Erfinder beauftrage ich hiermit den nachstehend benannten Patentanwalt (oder die nachstehend benannten Patentanwälte) und/oder Patent-Agenten mit der Verfolgung der vorliegenden Patentanmeldung sowie mit der Abwicklung aller damit verbundenen Geschäfte vor dem Patent- und Warenzeichenamt: (Name und Registrationsnummer anführen)

**POWER OF ATTORNEY:** As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. (list name and registration number)

Customer No.

And I hereby appoint

Telefongespräche bitte richten an:  
(Name und Telefonnummer)

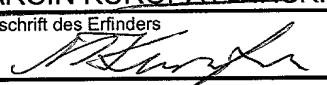
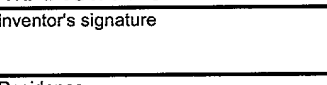
Direct Telephone Calls to: (name and telephone number)

Ext. \_\_\_\_\_

Postanschrift:

Send Correspondence to:

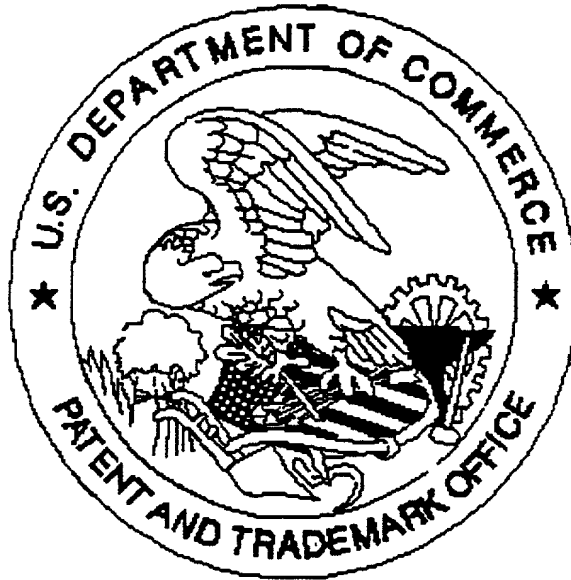
Bell, Boyd & Lloyd LLC  
70 West Madison Street, Suite 3300 60602-4207 Chicago, Illinois  
Telephone: +1 312 372 1121 and Facsimile +1 312 372 2098  
or  
**Customer No.**

Voller Name des einzigen oder ursprünglichen Erfinders. <b>MARCIN KUROPATWINSKI</b>	Full name of sole or first inventor: <b>MARCIN KUROPATWINSKI</b>
Unterschrift des Erfinders 	Inventor's signature 
Datum <b>30. Mai 01</b>	Date 
Wohnsitz <b>P-Gdansk, POLEN</b>	Residence <b>P-Gdansk, POLAND</b>
Staatsangehörigkeit <b>PL</b>	Citizenship <b>PL</b>
Postanschrift <b>c/o. ul. Danusi 1/3</b> <b>80434 P-Gdansk</b>	Post Office Address <b>c/o. ul. Danusi 1/3</b> <b>80434 P-Gdansk</b>
<div style="position: absolute; left: 200px; top: 0; font-size: 2em; font-family: cursive;">PLX</div>	
Voller Name des zweiten Miterfinders (falls zutreffend):	Full name of second joint inventor, if any:
Unterschrift des Erfinders	Second Inventor's signature
Datum	Date
Wohnsitz	Residence
Staatsangehörigkeit	Citizenship
Postanschrift	Post Office Address

(Bitte entsprechende Informationen und Unterschriften im Falle von dritten und weiteren Miterfindern angeben).

(Supply similar information and signature for third and subsequent joint inventors).

United States Patent & Trademark Office  
Office of Initial Patent Examination -- Scanning Division



Application deficiencies found during scanning:

☐ Page(s) \_\_\_\_\_ of \_\_\_\_\_ were not present  
for scanning. (Document title)

☐ Page(s) \_\_\_\_\_ of \_\_\_\_\_ were not present  
for scanning. (Document title)

☒ *Scanned copy is best available.*

*Drawings*